



# Universidad Autónoma del Estado de México

## Maestría en Ciencias de la Computación

---

Predicción de Tendencias Poblacionales Académicas en  
Municipios Conurbados del Estado de México Utilizando  
Inteligencia Artificial

---

Tesis que presenta

**Alfredo Alejandro Muñoz Morales**

Para obtener el Grado de

Maestro en Ciencias en Computación

Asesor de tesis:

**Dr. Jair Cervantes Canales**





# Autonomous University of Mexico State

## Computer Science Department

---

Prediction of Academic Population Trends in Conurbated  
Municipalities of the State of Mexico Using Artificial  
Intelligence

---

Submitted by

**Alfredo Alejandro Muñoz Morales**

As a fulfillment of the requirement for the degree of  
Master in Computer Science

Thesis Advisor

**Dr. Jair Cervantes Canales**

Texcoco City, Mexico State

December 2017



DICTÁMEN DE AUTORIZACIÓN Y OBTENCIÓN DE GRADO DE MAESTRÍA

Texcoco, México, a 20 de noviembre de 2017

Título del proyecto:

Predicción de Tendencias Poblacionales Académicas en Municipios Conurbados del Estado de México Utilizando Inteligencia Artificial

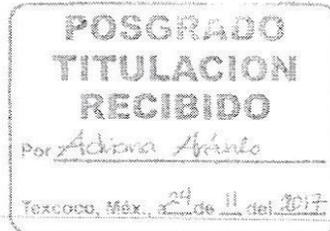
Tesista:

Alfredo Alejandro Muñoz Morales

Dictamen:

No. de revisión: 3

- Rechazado
- Sujeto a modificaciones
- Aceptado, condicionado
- Aceptado



Observaciones generales:

Aceptado para la impresión

Aceptado para la defensa de grado

Tutor Adjunto	Tutor Académico	Tutor Adjunto
Dr. en C. de la Comp. Farid García Lamont	Dr. en C. Jair Cervantes Canales	Dr. en I. S. José Sergio Ruiz Castilla



*Dedicated to my family, in every form.*

«*Nullius in verba.* »

Royal Society of London



# *Agradecimientos*

Gracias..



# Abstract

In society we can find many behaviors in the real world that we see around us, everything is in constant change or transformation, in constant movement, and all those activities and dynamic situations can be tried to represent by means of schemes. These schemes represent what we see in our daily life, and summarize in a simple way the complex things of our environment. That is called behavior patterns which are summarized or captured in mathematical models. In this thesis a study of the "Metropolitan Zone of the Valley of Mexico" is made. Currently, it is an area that has been in continuous growth and expanding more and more to other states as it is now Hidalgo. The main focus of the topic is to know the number of graduates of each educational level in the conurbated municipalities of the state of Mexico, from there it has as objective to study what social, demographic, financial, economic and other variables influence the number of graduates of each educational level, Pre-Primary, Primary, Secondary, Preparatory and Professional or technical, and to obtain from them forecasts for subsequent years to those of which we have information and thus in some way provide tools that are useful in the current take of decisions. The research focuses on the statistical study of graduates and causes that affect or influence the number of them in the conurbated municipalities of the state of Mexico.



# Resumen

En la sociedad podemos encontrar muchísimos comportamientos en todas las cosas que miramos en nuestro alrededor, todo está en constante cambio o transformación, en constante traslado o movimiento, y todas esas actividades y dinámicas situaciones se pueden tratar de representar por medio de esquemas, esquemas que representen aquello que miramos en nuestra vida cotidiana, y resuman en una forma sencilla las cosas complejas de nuestro entorno. A todo esto, anterior se le denomina patrones de comportamiento resumidos o plasmados en modelos matemáticos. En esta Tesis se hace un estudio de la “Zona Metropolitana del Valle de México”. Actualmente, es un área que ha ido en continuo crecimiento y expandiéndose cada vez más a otros estados como ahora lo es Hidalgo. El enfoque principal del tema es ahora conocer el número de egresados de cada nivel educativo en los municipios conurbados del estado de México específicamente, de allí se tiene como objetivo estudiar que variables sociales, demográficas financieras, económicas y demás áreas, influyen en el número de egresados de cada nivel educativo, Pre-Primaria, Primaria, Secundaria, Preparatoria y Profesional o técnico, y poder obtener de ellas pronósticos para años subsecuentes a aquellos de los que tenemos información y así de alguna manera brindar herramientas que sean útiles en la toma actual de decisiones. La investigación pues se centra en el estudio estadístico de egresados y causas que afectan o influyen en el número de los mismos en los municipios conurbados del estado de México.



# Índice general

<b>Resumen</b>	<b>VII</b>
<b>Resumen</b>	<b>IX</b>
<b>Índice general</b>	<b>XI</b>
<b>Índice de figuras</b>	<b>XV</b>
<b>Índice de cuadros</b>	<b>XIX</b>
<b>1. Introducción</b>	<b>1</b>
1.1. Problemática . . . . .	4
1.2. Justificación . . . . .	5
1.3. Objetivos y alcances del proyecto . . . . .	6
1.3.1. Objetivo General . . . . .	6
1.3.2. Objetivos Específicos . . . . .	6
1.4. Hipotesis . . . . .	7
1.5. Estado del arte . . . . .	7
<b>2. Preliminares</b>	<b>13</b>
2.1. Correlación . . . . .	14
2.2. Técnicas de regresión . . . . .	16
2.2.1. Regresión lineal . . . . .	16
2.2.2. Algoritmo Least-Mean-Square . . . . .	19
2.3. Neural networks . . . . .	21
2.3.1. Elementos de las RNA . . . . .	22
2.3.2. Función de costo y algoritmo de aprendizaje . . . . .	22
2.3.3. Modelo de RNA para predicción . . . . .	23
2.4. Support Vector Machines . . . . .	24
2.4.1. Teoría de aprendizaje estadístico . . . . .	24
2.4.2. Minimización del riesgo empírico . . . . .	25

2.4.3.	Dimensión Vapnik-Chervonenkis (VC)	26
2.4.4.	Minimización del Riesgo Estructural MRE	27
2.4.5.	Support Vector Machines	27
2.4.6.	SVM para datos linealmente separables	28
2.4.7.	SVM para datos linealmente no separables	29
2.4.8.	Kernels	29
2.4.9.	SVM para regresión (SVR)	30
2.4.10.	Least Square Support Vector Machine (LS-SVM)	31
2.5.	Selección de características utilizando un Algoritmo Genético	32
2.5.1.	Elementos de un algoritmo genético	33
2.5.2.	Algoritmo genético básico	34
	Población inicial	34
	Selección de individuos	35
	La Ruleta	35
	Sobrante Estocástico	35
	Cruza	36
	Mutación	36
	Condición de paro	36
2.6.	Técnicas de Evaluación de resultados	37
2.6.1.	Error de Predicción Medio	37
2.6.2.	Error de Porcentaje Absoluto Medio	38
2.6.3.	Error de Porcentaje Medio	38
2.6.4.	Error Cuadrático Medio	39
2.6.5.	Sumatoria del Error Cuadrático	40
2.6.6.	Error Cuadrático Medio Absoluto	40
2.6.7.	Raíz del Error Cuadrático Medio	40
2.6.8.	Error Cuadrático Medio Normalizado	41
<b>3.</b>	<b>Metodología</b>	<b>43</b>
3.1.	Conjunto de datos	43
3.1.1.	Variables utilizadas	44
3.1.2.	Municipios Obtenidos	46
3.2.	Pre-Procesamiento	46
3.2.1.	Limpieza	46
3.2.2.	Normalización de datos	47
3.3.	Cross-validation	47

3.4. Selección de características . . . . .	48
3.4.1. Reducción de la Dimensionalidad . . . . .	49
3.5. Predicción . . . . .	49
3.6. Análisis comparativo de resultados . . . . .	49
<b>4. Resultados experimentales</b>	<b>51</b>
4.1. Acolman . . . . .	53
4.2. Atizapan . . . . .	56
4.3. Chalco . . . . .	58
4.4. Chicoloapan . . . . .	60
4.5. Chimalhuacan . . . . .	64
4.6. Coacalco . . . . .	67
4.7. Cuautitlán de Romero Rubio . . . . .	69
4.8. Ecatepec . . . . .	71
4.9. Huixquilucan . . . . .	71
4.10. Ixtapaluca . . . . .	73
4.11. La Paz . . . . .	74
4.12. Naucalpan . . . . .	77
4.13. Nezahualcóyotl . . . . .	79
4.14. Nicolás Romero . . . . .	81
4.15. Tecamac . . . . .	83
4.16. Tepotzotlán . . . . .	85
4.17. Texcoco . . . . .	87
4.18. Tlalnepantla . . . . .	89
4.19. Tultitlan . . . . .	91
4.20. Análisis comparativo de los datos Reales VS Predicciones . . . . .	91
<b>5. Conclusiones</b>	<b>103</b>



# Índice de figuras

2.1. Correlacion a) Positiva b) Negativa c) Correlacion perfecta d) Sin correlacion . . . . .	15
2.2. Dispersión del cuarteto de Anscombe a) Positiva b) Negativa c) Sin correlacion d) Correlacion perfecta . . . . .	16
3.1. Metodología . . . . .	44
4.1. Resultados obtenidos utilizando Regresión lineal . . . . .	53
4.2. Resultados obtenidos utilizando redes neuronales . . . . .	54
4.3. Resultados obtenidos utilizando SVM con diferentes Kernels a) Gauss b) Polinomial c) RBF . . . . .	55
4.4. Resultados obtenidos utilizando diferentes métodos de regresion a) Regresión linealr b) Redes neuronales c) SVM método de Gauss d) SVM-Polinomial e) SVM-RBF . . . . .	57
4.5. Resultados obtenidos utilizando diferentes métodos de regresion a) Regresión linealr b) Redes neuronales c) SVM método de Gauss d) SVM-Polinomial e) SVM-RBF . . . . .	59
4.6. Resultados obtenidos utilizando diferentes métodos de regresion a) Regresión linealr b) Redes neuronales c) SVM método de Gauss d) SVM-Polinomial e) SVM-RBF . . . . .	61
4.7. Resultados obtenidos utilizando diferentes métodos de regresion a) Regresión linealr b) Redes neuronales c) SVM método de Gauss d) SVM-Polinomial e) SVM-RBF . . . . .	63
4.8. Resultados obtenidos utilizando diferentes métodos de regresion a) Regresión linealr b) Redes neuronales c) SVM método de Gauss d) SVM-Polinomial e) SVM-RBF . . . . .	65
4.9. Resultados obtenidos utilizando diferentes métodos de regresion a) Regresión linealr b) Redes neuronales c) SVM método de Gauss d) SVM-Polinomial e) SVM-RBF . . . . .	66

4.10. Resultados obtenidos utilizando diferentes métodos de regresion a) Regresión linealr b) Redes neuronales c) SVM método de Gauss d) SVM-Polinomial e) SVM-RBF . . . . .	68
4.11. Resultados obtenidos utilizando diferentes métodos de regresion a) Regresión linealr b) Redes neuronales c) SVM método de Gauss d) SVM-Polinomial e) SVM-RBF . . . . .	70
4.12. Resultados obtenidos utilizando diferentes métodos de regresion a) Regresión linealr b) Redes neuronales c) SVM método de Gauss d) SVM-Polinomial e) SVM-RBF . . . . .	72
4.13. Resultados obtenidos utilizando diferentes métodos de regresion a) Regresión linealr b) Redes neuronales c) SVM método de Gauss d) SVM-Polinomial e) SVM-RBF . . . . .	75
4.14. Resultados obtenidos utilizando diferentes métodos de regresion a) Regresión linealr b) Redes neuronales c) SVM método de Gauss d) SVM-Polinomial e) SVM-RBF . . . . .	76
4.15. Resultados obtenidos utilizando diferentes métodos de regresion a) Regresión linealr b) Redes neuronales c) SVM método de Gauss d) SVM-Polinomial e) SVM-RBF . . . . .	78
4.16. Resultados obtenidos utilizando diferentes métodos de regresion a) Regresión linealr b) Redes neuronales c) SVM método de Gauss d) SVM-Polinomial e) SVM-RBF . . . . .	80
4.17. Resultados obtenidos utilizando diferentes métodos de regresion a) Regresión linealr b) Redes neuronales c) SVM método de Gauss d) SVM-Polinomial e) SVM-RBF . . . . .	82
4.18. Resultados obtenidos utilizando diferentes métodos de regresion a) Regresión linealr b) Redes neuronales c) SVM método de Gauss d) SVM-Polinomial e) SVM-RBF . . . . .	84
4.19. Resultados obtenidos utilizando diferentes métodos de regresion a) Regresión linealr b) Redes neuronales c) SVM método de Gauss d) SVM-Polinomial e) SVM-RBF . . . . .	86
4.20. Resultados obtenidos utilizando diferentes métodos de regresion a) Regresión linealr b) Redes neuronales c) SVM método de Gauss d) SVM-Polinomial e) SVM-RBF . . . . .	88

4.21. Resultados obtenidos utilizando diferentes métodos de regresion a) Re- gresión linealr b) Redes neuronales c) SVM método de Gauss d) SVM- Polinomial e) SVM-RBF . . . . .	90
4.22. Predicciones obtenidas para Pre-Primaria . . . . .	93
4.23. Predicciones obtenidas para Primaria . . . . .	95
4.24. Predicciones obtenidas para Secundaria . . . . .	96
4.25. Predicciones obtenidas para Bachillerato . . . . .	98
4.26. Predicciones obtenidas para Profesional . . . . .	99



# Índice de cuadros

2.1. Expresiones que se utilizan en la genética con su estructura equivalente en un algoritmo genético . . . . .	33
4.1. Valores reales conjunto de datos Tlanepantla . . . . .	92
4.2. Comparativa de datos reales y resultados utilizando las diferentes técnicas de regresión . . . . .	93
4.3. Métricas de comparación para el municipio de Tlanepantla (Pre-Primaria)	94
4.4. Métricas de comparación para el municipio de Tlanepantla (Primaria) .	96
4.5. Métricas de comparación para el municipio de Tlanepantla (Secundaria)	97
4.6. Métricas de comparación para el municipio de Tlanepantla (Bachillerato)	99
4.7. Mejores resultados de las técnicas utilizadas . . . . .	100
4.8. Métricas de comparación para el municipio de Tlanepantla (Profesional)	101



# Capítulo 1

## Introducción

Dentro de nuestras actividades cotidianas encontramos comportamientos o actividades que continuamente están aconteciendo, es importante ver que nuestras vidas están llenas de diferentes contextos, actividades, gente, comportamientos, reacciones y demás, y si prestamos atención y somos más observadores lograremos ver algunos de estos comportamientos o actividades, encontrando en ellos muchos que se repiten una y otra vez, y lo hacen bajo ciertas condiciones, es decir que a ciertos fenómenos u acontecimientos se presenta una respuesta (inmediata o no) de parte de la sociedad o por lo menos de la mayoría de la gente, tenemos varios ejemplos sencillos, uno lo podemos mirar el siguiente evento: “Ausencia de gente en los cibercafé cuando hay lluvia”, de lo cual abundamos, cuando no llueve hay más probabilidad de encontrar gente en los cibercafé, y en general en los comercios exteriores, pero cuando llueve este aglutinamiento de gente disminuye, en este ejemplo la causa es lógica, la gente no sale porque hay lluvia que provocara a su vez incomodidad en los que resulten mojados por ella; La causa sería la lluvia, la consecuencia sería la ausencia o disminución de clientes en el cibercafé. Un ejemplo más podría ser la múltiple venta de líquidos en días donde hay algún juego de fútbol importante, cuando hay partido aumentan la proporción de venta de líquidos que en el resto de los días donde no hay partido, por lo tanto, los vendedores se prepararan para el evento y poder satisfacer la demanda de líquidos para toda la gente que los consumirá, a más de las ganancias que ellos saben, tendrán, también encontramos una respuesta lógica en este ejemplo pues los días de juego generalmente son días soleados y de alta actividad física aun para los espectadores, la tensión de juego, la prisa de llegar al partido, la euforia de un gol anotado, etc. . . tantas actividades que provocan la deshidratación de la persona y generan una necesidad en ella de consumir líquidos, causas y consecuencias. Como estos, otros muchos comportamientos que son fáciles de identificar, no nos es difícil entender su naturaleza ni explicación, sin embargo, existen otros eventos cuya explicación es menos lógica, y la mayoría de

veces no somos capaces de identificarlas. Son fenómenos que son cotidianos para nosotros pero que sin un momento de análisis (en ocasiones profundo), no logramos identificar, por ejemplo, tendríamos cuando vemos que la luz del semáforo se pone en amarillo los conductores aceleran para lograr cruzar antes que cambie a rojo, esta tan implícito en nuestra vida que se vuelve difícil distinguirlo en ocasiones.

Como este comportamiento hay muchísimos más, algunos a nivel muy personal, como los hábitos de una familia en particular, como el hábito de ir a comer cada martes a la pizzería porque tiene 2 por 1, también hay hábitos colectivos, un ejemplo en las personas que vemos cada domingo desfilar por las calles con sus playeras de fútbol cuando su equipo favorito ha ganado.

Dijimos que muchos comportamientos se pueden lograr identificar solamente con nuestra concentración, aunque también hay muchos otros que no son fácilmente identificables, de hecho, tenemos la mayoría que son difícilmente visibles pues son patrones que se reflejan masivamente y solo se pueden encontrar con estudios estadísticos adecuados, es nuestro deseo encontrar más de ellos. Muchos patrones en la sociedad son los responsables de cambios o fenómenos masivos que son los que dan forma a una sociedad o cultura, no todos tienen la misma influencia o fuerza de cambio en la gente, sin embargo, es posible medirlos, conocerlos, estudiarlos, y tendremos la posibilidad de controlar algunos de ellos en pro de la gente.

Una de las formas más aplicada en la actualidad es la mercadotecnia y publicidad, pueden lograr que la gente haga o compre lo que una empresa televisiva desea, y todo a través de imágenes y sonidos, es un uso lucrativo y destructivo a la vez por la forma en que se ocupa, sin embargo, es posible incrementar dichos estudios sobre la gente para poder diseñar planes que irían desde el corto al largo plazo, aunque más al mediano y largo plazo y podremos ocupar en proyectos sociales de desarrollo cultural y demás.

Hemos estado hablando de comportamientos colectivos, sin embargo, hay comportamientos todavía más globales, ejemplificando de nuevo tenemos la forma de comprar en un mercado, la marca favorita de un refresco según el país, el nivel de estudios promedio de una población, la tendencia de la gente de una región geográfica a votar por determinado partido, o aún más masivo, la reacción de una persona al ser agredida, la reacción al sentirnos en peligro que en realidad deja de ser una variable sociocultural para convertirse en una variable humana. Entonces entraremos que hay variables y/o comportamientos que se desarrollan con base en una cierta sociedad o cultura, pero que se van haciendo cada vez más generales, para pasar de ser particulares de una sociedad a ser globales del género humano, la tendencia cambia según las causas que

produzcan los efectos o patrones, dejan de ser específicos para ser generales, aun en un sentido opuesto pueden particularizarse tanto como deseemos, los comportamientos un municipio o estado de la República Mexicana, no serán igual al resto de la sociedad mexicana, aunque nuestra cultura sea sumamente similar entre nosotros y completamente diferente a la de gente en países europeos, aunque también el comportamiento de una familia puede variar muchísimo del comportamiento de otras familias dentro de dicho municipio o estado aunque estén allí dentro.

Existen características globales y particulares en muchísimos niveles, podríamos concluir la idea con decir que los patrones de comportamiento de una sola persona variarían mucho de otra que viva en su misma casa, citaremos el proverbio popular “cada cabeza es un mundo”. Sin embargo, por mucho que pueda cambiar los patrones de comportamiento, existe una forma de estudiarlos y englobarlos de una manera general, no con ello decimos que tendremos total comprensión de dichos patrones, pero si analizar por métodos de frecuencia cuales son más probables según las condiciones.

A todos los comportamientos que surgen como respuesta a un acontecimiento específico, o que tal vez se repiten una y otra vez estadísticamente se les denomina patrones, a todos los diferentes razones que nos permiten comportarnos como lo hacemos se les denomina variables, otro ejemplo sería, mientras hay gente que daría todo su dinero a una casa de ayuda humanitaria también hay quien quitaría todo el dinero que un tercero haya dado a dicha causa de ayuda humanitaria, y la razón puede solo basarse en la forma en que la persona vivió su niñez, o su juventud, ¿cómo podría existir una variable con tanto tempo que logre influir al en mi vida al día de hoy? Bueno, es perfectamente posible, podemos decir que somos el resultado de una confluencia impresionante de variables que se han cruzado a lo largo de nuestra vida, no todas las variables tienen una repercusión inmediata, ni influyen todas con la misma fuerza en nuestra vida, pero estamos seguros que influyen de una u otra forma, el estudio nos llevara a medir todo lo que quede en nuestro alcance y sea accesible. Entonces una variable, como su nombre lo dice, un aspecto específico que se puede medir pero que puede cambiar, no se repite en todos de igual forma y cambia según su fuente.

La relación variables – patrones es sumamente interesante, se convierte en un aspecto de causa – efecto, ya que a diferentes variables tenderemos diferentes patrones, en el estudio específico de esta investigación se desea conocer solo un aspecto particular de la sociedad, pues estudiar todas las variables dentro de un área requeriría mayor inversión de recursos, se seleccionó un sector social de amplio interés por lo tanto, educación, se desarrolla específicamente bajo el ámbito “nivel académico”, de amplio interés

actual y profunda aplicación; Se desea desarrollar un modelo que permita obtener las variables que influyen en el nivel académico alcanzado por la gente en los municipios conurbados del Valle de México, para conocer qué razones los lleva a lograr su grado académico, así como el área en que se desarrollan, podemos identificar mayor fuerza de ciertas ramas científicas según la región.

La regio de municipios conurbados del Valle de México, alberga la mayor población dentro del estado, su densidad es mucho mayor, como también lo es su diversidad cultural respecto a otros municipios del Estado de México, posee gente inmigrante de otros estados que en su tiempo buscaban mejores oportunidades de empleo, por lo que la diversidad racial también crece, son municipios con un alta actividad industrial y económica, muchos de ellos aportaron en otros años gran cantidad de producción en el país, sin embargo hoy en día varios de ellos están ahora en estado de quiebra. Actualmente la zona metropolitana del Valle de Toluca posee un mayor porcentaje de aportación a la producción nacional. Debido a su crecimiento repentino en otros años, varios de estos municipios no cuentan con una planeación y desarrollo urbano adecuada, en municipios como Naucalpan el crecimiento fue tan repentino que se perdió total estructura en el trazo de las calles y bloques. Es entonces la zona metropolitana del Valle de México con los municipios que pertenecen al Estado, una zona de amplio interés estadístico, pues diverge totalmente de los aspectos socioculturales con el resto de municipios en el estado, la mayoría de lado oeste del mismo, circunvecinos de la zona Toluca.

## 1.1. Problemática

Dentro de la sociedad hay miles de millones de personas que viven en muy diversas zonas geográficas, cada grupo de personas tienen características diferentes, que se definen en diferentes variables, a algunas las denominamos culturales, otras sociales, económicas, demográficas, entre muchas más.

Está demostrado que el nivel académico de una persona, que también puede denominarse una variable social, ofrece a la persona una mejor oportunidad de incrementar su calidad de vida, que no solo incluye el aspecto económico sino muchos más como el laboral, de seguridad social, salud, cultura, entre otros, todo ello en conjunto son capaces de otorgar a la persona una mejor calidad de vida a la persona y su familia.

Sin embargo, dado que cada sociedad y cultura difieren entre sí, el acceso a la educación suele ser limitado en algunas de ellas, en muchas ocasiones desde nivel preescolar,

en otras se dificultan los niveles medios y de manera más común en niveles profesionales; Las razones por las que se vuelve difícil incrementar el nivel académico varían mucho, a tal grado que pueden ir desde falta de solvencia económica (que obliga a la persona a priorizar su gasto en alimentación en vez del gasto en educación), hasta la simple falta de interés por asistir a la escuela, las razones pueden ser muchísimas.

También es muy importante saber que lo antes mencionado es de gran preocupación para el gobierno de un país, pues la población con mejor nivel educativo puede aportar más a su organización social, puede desempeñar más trabajos de calidad, por ejemplo, también puede ser de más confianza para empresas que son los principales inversionistas de una economía, de una persona con buen nivel académico se puede esperar un delincuente menos, entre otros ejemplos. A los gobiernos les beneficia más tener gente con alto nivel académico, sin embargo, como vimos antes poseer un nivel académico alto es una cuestión de convergencia de varias variables.

¿Los gobiernos tienen a su disposición o posibilidad el manejo de muchas de las variables en su región, así como de otras no, sin embargo, como saber que variables son las que les convendría estimular para fortalecer la posibilidad de desempeñar mejores niveles académicos en su población? ¿Como? Y en cuanto alterar una determinada variable, cuál sería su costo, cuanto tiempo tomaría, el gobierno sabe que un nivel académico alto para la gente no se adquiere en un año, es un proyecto a mediano plazo, pero cuyas repercusiones tienen efectos de largo plazo en beneficio de la sociedad.

Es posible desarrollar un modelo que representa dichas variables e identifique los principales aspectos a considerarse dentro del trabajo que se debe hacer en pro de la educación, como estrategia de crecimiento a largo plazo.

## 1.2. Justificación

Si bien existen multitud de herramientas para medición de variables, y modelos clásicos o frecuentistas para contabilizarlas, no es fácil encontrar, sobre todo en México, métodos basados en Inteligencia Artificial sobre estudios de población.

La estadística es abundantemente usada en los estudios demográficos, sin embargo, ha quedado limitada en los estudios más recientes debido a que los modelos que se utilizan para evaluarlas y trabajarlas son modelos estadísticos que se encuentran basados en estadística clásica o frecuentista, y que las variables socio - demográficas no necesariamente se ajustan a estos modelos, por el contrario, los datos que se recopilan

son reales, y por sí mismo no se ajustan - en la mayoría de los casos - a alguna función normal o cualquier otra, que con frecuencia son utilizadas para el modelado de la distribución de los datos. El trabajo de la Inteligencia Artificial es generar modelos de evolución estructurada que permiten la simulación de algún basado en el proceso de aprendizaje humano y que nos da la posibilidad de analizar datos en una forma más pura y real de las variables.

Tener la posibilidad de ocupar esta tecnología estadística en los estudios sociodemográficos nos permitirá tener información, que por su complejidad y estructura hasta hoy ha sido un poco complicado de analizar o interpretar adecuadamente en su totalidad, por lo tanto este trabajo es de importancia considerable al ofrecer una solución o alternativa de análisis y estudio de variables que tengan que ver o influyan con la falta de posibilidad de adquisición de un mejor nivel académico, las variables que influyen también de manera indirecta, y las que influyen de manera directa.

Si en la actualidad existen modelos que nos puedan proporcionar una mejor respuesta a los problemas que se plantean, una mejor solución o resultado que sea más veraz y preciso al generar sus resultados. Una herramienta que de la posibilidad de trabajar con los datos aprovechando más la información que poseen, entonces conviene asirnos de las herramientas que nos generen estas facilidades o ventajas, es la forma en que nos ayudarían la herramienta que nos da la Inteligencia Artificial.

## **1.3. Objetivos y alcances del proyecto**

### **1.3.1. Objetivo General**

Construir un modelo basado en Inteligencia Artificial que permite predecir el número de egresados por nivel educativo de una población específica.

### **1.3.2. Objetivos Específicos**

1. Construir un modelo basado en Inteligencia Artificial específicamente Redes Neuronales Artificiales, que permite predecir los egresado por nivel educativo de una población específica.
2. Determinar de los patrones encontrados, los referentes a los aspectos que se vean involucrados en el objeto de estudio que es Nivel Académico.

3. Poder realizar la identificación de las variables determinantes en el estudio y análisis de Nivel Académico en la población.

## 1.4. Hipotesis

1. Es posible realizar utilizando técnicas de Inteligencia Artificial la selección de mejores características de variables en general que puedan proporcionar información determinante en la predicción de cantidad de egresados de los diferentes niveles educativos
2. Es posible plantear modelos basados en Inteligencia Artificial que permitan realizar predicciones certeras acerca de los valores que podría tener en un futuro el número de egresados de los diferentes niveles educativos.
3. Es posible realizar un análisis utilizando métricas de evaluación de error para determinar qué modelo de Inteligencia Artificial es el mejor en la predicción y ajuste al número de egresados de los diferentes niveles educativos.

## 1.5. Estado del arte

La presente investigación se concreta al estudio de diferentes modelos de predicción enfocados específicamente al análisis del número de egresados de los diferentes niveles académicos desde Pre Primaria hasta Profesional o Técnico, el estudio actual se realiza con la comparación de los modelos los modelos usados son clasificables como modelos de Inteligencia Artificial, modelos estadísticos que pueden funcionar como clasificadores o como predictores, generalmente cuentan con la cualidad de recibir retroalimentación con el paso de los experimentos. Los modelos que se ocupan en esta investigación de IA son Redes Neuronales Artificiales y Maquinas de Vectores Soporte, también se ocuparan modelos de Regresión Lineal, aunque no están considerados como modelos de Inteligencia Artificial, se realiza el presente preámbulo para hablar de los antecedentes y más modernas investigaciones sobre el área. Las RNA han sido ya de hace años una importante técnica de IA, a la que se recurre como modelo de predicción, ha ofrecido excelentes resultados, existen diferentes modelos dentro de las RNA que se han ido desarrollando con el paso del tiempo, un lapso de tiempo tuvo mucho auge, pero fue resurgiendo con la llegad de nuevos modelos de RNA. La forma que en la se hará la comparación de modelos es ocupando la información del número

de egresados en los grados de estudio y esto incluye estudios de la población, que nos ayuda en los análisis de cómo se comporta a lo largo del tiempo y en los diferentes lugares, a distintas edades y demás variables. Actualmente lo anteriormente dicho tiene mejoras y actualizaciones, que se siguen realizando de entre las diferentes áreas que se llevan a cabo, desde nuevos modelos hasta nuevas investigaciones del área. Existen muchos modelos, actualmente hay varias formas de realizar el pronóstico de una cierta investigación, cada modelo aporta ventajas y desventajas. Haiyin. L y Zhiqi. L., nos muestran un modelo de predicción de tamaño de población que está basado en ecuaciones diferenciales, métodos de regresión y recurrencia, que ha arrojado buenos resultados, pero que no ha sido modelado bajo modelos actuales de IA. Debido a que las ecuaciones diferenciales se prestan para modelar procesos con cambios se han adecuado para poder simular, a través de los resultados, las predicciones en el crecimiento de la población de China. Los países de población grande, variable, o que presentan problemas de población suelen ser los más estudiados en el caso de crecimiento o decaimiento en su población, China es uno de ellos, por ello se buscan diferentes modelos con precisión puedan dar un pronóstico de crecimiento en dichas naciones, ya sea por población alta de la que se desea pueda reducirse en un plazo mediano, como también naciones que desean su población pueda incrementar por los problemas que la caída de la misma pueda afectar en el futuro las situaciones del país, tal es el caso de Rusia. Mientras Haiyin. L y Zhiqi. L. buscan un modelo que les ayude en la estimación de la población por causa de la sobrepoblación de China y que a la larga representara un problema grave para su país, las predicciones buscan otorgar herramientas que ayuden a la mejor toma de decisiones, sobre todo del gobierno de dicha nación, y que produzcan a la larga resultados que desemboquen en reducción de su población. Los estudios de Población se han vuelto cada vez más populares y necesarios al momento de tomar decisiones, si nosotros consideramos el crecimiento de la población en todas partes del mundo, viene a generar situaciones tanto favorables como desfavorables en los países, por ejemplo, tenemos los países donde las pirámides poblacionales muestran una población avejentada que da pocas probabilidades de escenarios positivos en el futuro, su población se está quedando sin jóvenes quienes a su vez representan la fuerza laboral, estudiantil, mano de obra, y demás características importantes en el desarrollo de una nación, varios de estos países son considerados primermundistas, sin embargo hoy presentan un déficit juvenil en sus pirámides poblacionales. De allí la demografía se ha vuelto una rama importante en el estudio y cuidado de los países, pues de ella dependerá el futuro de los mismos, de nada sirve un país rico, que a la larga no tenga

gente que pueda disfrutar o continuar con la riqueza actual. Sin embargo tenemos otro caso contrario, el caso donde la población tan numerosa amenaza la seguridad y bienestar del país, a nivel mundial se acrecienta día a día la población en muchos países del mundo (aunque en otros mengua) hace más de 200 años Thomas Malthus publicó un panfleto donde decía que la población en el mundo siempre tendería a crecer, hoy la población a nivel mundial ha crecido más de siete veces desde entonces, según Guo D., Zhang Y., He L., Zhai K., y Tan H., quienes hacen un investigación acerca de un modelo que se ajuste y pronostique con buena precisión la población mundial. Ellos argumentan que, en muchas investigaciones sobre población, se han ocupado variables propias de la misma como mortalidad, fecundidad, u otros factores; También se han empleado factores como las políticas públicas, el entorno natural, economía, cultura, entre otras opciones, sin embargo, ellos proponen un método en el que se ocupan los datos en bruto de la población de hace 10,000 años, esto es, 8,000 años a.C., pensando en el asunto de que en una variable a la que se le debe sacar el mayor provecho posible. Las guerras o desastres naturales también han sido un factor determinante en el número de gente, según la investigación que realizaron, el modelo que ocupan ellos es una red neuronal artificial (RNA) activada por un grupo de Polinomios Chebychev clase 1 y construida bajo el algoritmo “weights and structure determination” (WASD) para su mejor desempeño. Con esto nos damos cuenta de que el resultado deseado es información necesaria para la toma de decisiones hoy en día, ya que mientras el planeta tenga un mayor incremento cada vez de su población la situación del sustento y demás cosas necesarias para el hombre se vuelven una preocupación para los gobiernos, las empresas públicas y privadas, y otras entidades de salud, economía y demás factores determinantes en la vida del hombre en la tierra. Los resultados que los obtuvieron en el ajuste a los datos históricos de la población humana han sido muy buenos, determinando un modelo que se ajusta adecuadamente al crecimiento en los últimos 10,000 años, el experimento se realizó hasta el año 2013, ofreciendo un modelo preciso en el modelado de la población humana hasta ese año. En cuanto a la predicción se generó un modelo con la misma red, los resultados son variados dado el mínimo local que se utilice en el algoritmo WASD sin embargo los resultados con menor nivel de error asumen que en la siguiente década las estimaciones son crecientes hasta un máximo en el 2020 tras el cual comenzara una caída de la población a nivel mundial. También los estudios se pueden dar para un determinado grupo de gente, no necesariamente un país o el mundo entero, en China existe un grupo étnico denominado “Wa”, según el censo del mismo país, dicha población ha ido incrementando, nos cuentan Mei. J. et. Al., en su estudio de proyección de

crecimiento de la población en dicho grupo étnico al 2060 ellos encuentran dificultades en la información existente acerca de la población Wa, ya que es un grupo no documentado en algunos periodos de tiempo en la antigua china, o en la china contemporánea donde se les discriminaba, la gente de Wa no declaraba su identidad por temor, siendo un grupo étnico no se le dio antes importancia a la contabilización de su población aún menos tras guerras en el país, que generaban disturbios. La población de Wa vive entre los límites de China, Laos, Birmania y Tailandia, por esta misma razón, en los censos realizados en China, que se realizan cada 10 años, solo se tomó cuenta de la población dentro del territorio nacional, por lo tanto, la información sobre el resto de pobladores fuera del país, se obtuvo mediante información proporcionada por los jefes de dicha etnia, discrepando bastante la información que ellos proporcionaban de la información real de pobladores. Todo esto reputa fuertemente al momento de querer realizar predicciones de la población, algunos modelos estadísticos tienen fortalezas al momento de realizar proyecciones y utilizando poca información histórica, otros requieren de muchos datos para poder realizar sus predicciones, en este caso, ellos ocuparon un software de proyección "Spectrum". Los censos en China se realizan periódicamente cada 10 años desde 1990, de allí hacia atrás los tres censos previos no se llevaron a cabo en intervalos equidistantes de tiempo, la información que se tiene de cada censo no es por lo tanto equidistante, así mismo en aquel tiempo se estima que una buena parte de pobladores Wa no declaro que lo era. Mei. J. et. Al., trabajaron con tres esquemas diferentes, alto, medio y bajo. Según el Software que ocuparon, los resultados para la predicción de cada quinquenio por varios periodos después del 2010 y hasta el 2060, arroja una curva cuyo máximo alcanza más de 458 mil habitantes Wa en 2027 para después volver a decaer, esto con el "esquema bajo", con el "esquema medio" se alcanzan más de 490 mil en 2039 para caer después de ese año. Los factores que tomaron en la determinación de la población fueron primer año de población, tasa de fertilidad según el año base, la estructura de edades y el patrón de reproducción de las mujeres en edad reproductiva, predicción de la tasa de sexo en los nacimientos, predicción de la esperanza de vida, selección de modelo de tabla de mortalidad, y la migración. Variables que fueron utilizadas en la predicción de crecimiento de población Wa en los siguientes años. Según ellos con el tercer esquema "alto" proyectan el máximo en el año 2060 con más de 559 mil habitantes. Los estudios de población pueden variar muchísimo, y cada uno presenta dificultades a la hora de hacer el acomodo de variables independientes, ciertamente hay algunas más significativas que otras, dicho de otra manera, variables

que aportan más información útil para las proyecciones que otras. AL final de su estudio concluyen en que sus resultados están sujetos a una gran cantidad de variables no predecibles, como las políticas de gobierno, así pues, la proyección queda en espera de otro posible cambio en los resultados, según Mei. J. et. Al. el factor más influyente en algún cambio será la migración en dicha población. Del mismo modo se tiene que países con las tasas de crecimiento tan vigiladas como China y Rusia buscan mejores formas de conocer su población futura, ya sea como mencionamos antes por escasez o exceso de población, como son los casos de Rusia y China respectivamente. Como hemos mencionado los modelos sirven en varias situaciones, por ejemplo, Rusia presenta un problema de decaimiento en su población, esta situación compromete el bienestar futuro del país, una población decreciente es un riesgo para el futuro de una nación, en el caso de Rusia se desea el incremento de la población, los modelos de estimación de población buscan variables que ayuden a llevar a cabo un cambio favorable en la tasa de crecimiento. Por el contrario, en el caso de China la población excesiva es una amenaza también para el futuro del país, ya que hay variables que se verán perjudicadas si las tasas de crecimiento no cesan y por el contrario se invierten. Zhang Y., et. al. (2014) en el caso de Rusia realizaron una investigación ocupando una RNA de tres capas, una de ellas oculta, aprovechando los beneficios del algoritmo “weights and structure determination” (WASD) que consiste en poder determinar los pesos con los que se comunicaran las neuronas de una capa a otra adecuadamente, así como calcular el número óptimo de neuronas que la red deberá ocupar para su desempeño y mejor predicción. El algoritmo ha sido ocupado recientemente con éxito en los modelos de RNA, Zhang Y., et. al. (2015) realizaron también un modelo de ARN con WASD aplicado a la población de China, para poder proyectar su tendencia, en este caso se desea saber que población habrá por motivo de que es muy alta la población que existe en China y así como con Rusia, aunque en un sentido opuesto en la población, esta situación podría concluir en problemas futuros para el país. Ambos estudios buscan generar tendencias poblacionales más ajustadas a la realidad pasada y actual, ya que información precisa en la actualidad podría ser clave en las decisiones que el país tome, así como también instituciones que sean de destacada importancia en dicho país. Los algoritmos WASD aplicados a las redes utilizadas, optimizan sus resultados permitiendo la obtención de datos mucho más precisos, si ya de por si un modelo de RNA es preciso, este algoritmo permite que los resultados sean basados en mínimos errores. En el estudio realizado para China podemos encontrar que se deseaba realizar una mejor aproximación a la

información real que el modelo con el que se hicieron las comparaciones de rendimiento llamado World Population Prospects (WPP) revisión 2012, los resultados obtenidos fueron según Zhang Y., et. al. (2015) más precisos en su modelo RNA-WASD que en el WPP. El asunto en que difieren los dos modelos también se caracteriza por las variables que se utilizan, el modelo de estimación que usa WPP se basa en la suma y resta de elementos como el número de nacimientos, de inmigrantes, de defunciones y de emigraciones, mientras que la red neuronal utiliza únicamente el número de habitantes de cada año. Por esta causa es destacable mencionar que un modelo de inteligencia artificial en este caso un RNA logro más precisión o ajuste a los datos reales que el modelo de estimación usado en el WPP. En el caso del modelo aplicado a Rusia que podemos decir es muy parecido al de China, utilizo datos de 2013 años de antigüedad, se estimó la población del año 1000 d.C. al 1800 d.C. y se realizó la corrección hacia 1897 y de allí se indicaron varias posibilidades en el crecimiento de la población de Rusia, según el punto mínimo de error en el algoritmo WASD la población rusa tiene más probabilidades de seguir decreciendo en los años futuros que de aumentar, de allí que se tendrá problema en el futuro por falta de población. También tenemos que en este mismo intento de lograr una mejor predicción se han realizado también previamente acerca de la población de los países, la comparación de modelos en busca del mejor o que más se ajuste a los datos reales, así como el que mejor se ajuste a las predicciones necesarias para la toma de decisiones. Múltiples ocasiones la comparación de modelos ayuda en estos casos a saber cuáles son los mejores modelos que se pueden implementar bajo determinadas circunstancias, en el caso de la población los modelos van desde regresiones lineales hasta modelos avanzados de Inteligencia Artificial (AI) basados en estadística no paramétrica, Poblacionalmente se desea conocer diversos y abundantes datos sobre el pasado y futuro, permitiéndonos aplicar modelos o diseñar planes de desarrollo adecuados basados en información lo más cercana a la realidad.

# Capítulo 2

## Preliminares

La predicción de series de tiempo ha sido una área de investigación activa las pasadas décadas. La predicción exitosa de series de tiempo es un reto muy importante en diversas áreas de investigación desde economía, planeación de negocios, control de inventarios y producción, predicción de tiempo, control y procesamiento de señales, etc. A lo largo de los últimos años varios modelos de predicción han sido desarrollados. Las técnicas de predicción incluyen enfoques basados en regresión. Estas técnicas de predicción utilizan datos históricos y exploran los patrones para crear un modelo de predicción.

En el pasado, métodos estadísticos convencionales eran utilizados para predicción de series de tiempo. Sin embargo, la predicción de series de tiempo es a menudo totalmente no lineal e irregular. Es por ello que varios métodos de predicción no lineales han sido propuestos en la literatura actual. La precisión de la predicción de series de tiempo es fundamental para tomar decisiones certeras en muchos de los procesos de diversos campos de aplicación, debido a ello, la investigación de modelos que mejoren la efectividad en la predicción ha sido una constante en los últimos años.

En esta Sección se da una breve introducción a los métodos de regresión utilizados en esta tesis, así como las técnicas de selección de características y técnicas de validación de resultados utilizadas. Los modelos que se utilizaron en la investigación son cinco, cada uno tiene diferentes propiedades, dichos modelos se han tomado para revisar su rendimiento o ajuste al comportamiento que muestran los datos de los diferentes municipios, los modelos son: Modelo de Regresión lineal, Redes Neuronales Artificiales, Máquinas de Vectores de Soporte con kernel Radial, kernel Gaussiano y kernel Polinomial, se analiza en esta Sección cada uno de forma independiente y detallada.

## 2.1. Correlación

Antes de intentar ajustar un modelo lineal a los datos observados, primero es necesario evaluar si existe o no una relación entre las variables de interés. Esto no necesariamente significa que una variable causa otra, sino que existe una relación significativa entre estas dos variables. Un diagrama de dispersión puede ayudar a determinar si existe una fuerte relación entre dos variables. Si no existe una asociación entre las variables dependiente e independientes el modelo de regresión lineal ajustado a los datos proporcionados no será útil. Una métrica importante para calcular la asociación entre dos variables es el coeficiente de correlación.

La correlación determina la relación o dependencia que existe entre dos variables que intervienen en una distribución bidimensional. Si una de las variables influye en los cambios de otra, se dice que las variables están correlacionadas o que existe una correlación entre ellas.

El coeficiente de correlación lineal se expresa mediante la ecuación siguiente:

$$r = \frac{\sigma_{xy}}{\sigma_x \sigma_y} \quad (2.1)$$

$$= \frac{E[X - \mu_x][Y - \mu_y]}{\sigma_x \sigma_y} \quad (2.2)$$

El coeficiente de correlación lineal es algunas veces referido como el coeficiente de correlación de Pearson, en honor a quien lo desarrolló Karl Pearson. El valor de  $r$  está en el intervalo  $-1 \leq r \leq +1$ . Los signos  $+$  and  $-$  son utilizados para correlación positiva y correlación negativa respectivamente. La Figura 2.1 muestra imágenes con datos que poseen correlación positiva, correlación negativa, sin correlación y correlación perfecta.

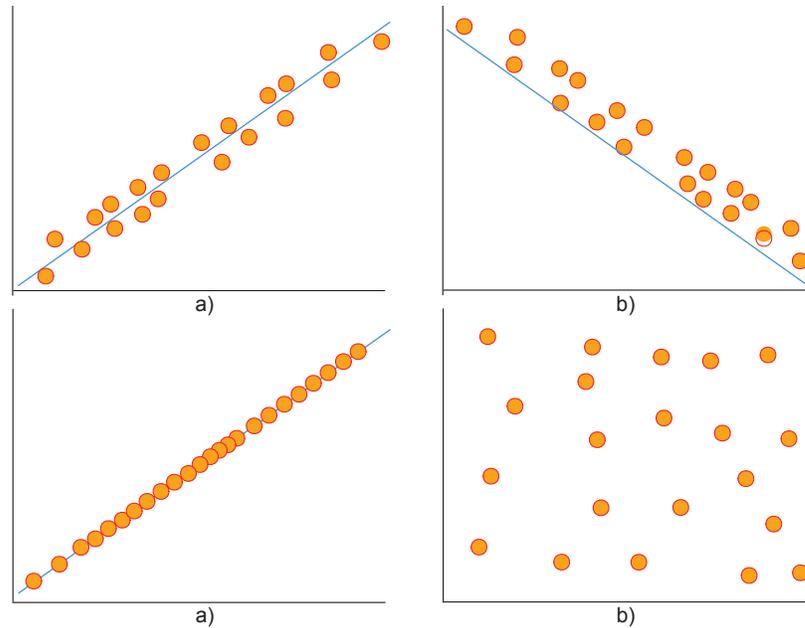


FIGURA 2.1: Correlacion a) Positiva b) Negativa c) Correlacion perfecta d) Sin correlacion

El coeficiente de correlación indica la fuerza de una relación lineal entre dos variables, pero generalmente no caracteriza completamente su relación. En particular, si la media condicional de  $Y$  dada  $X$ , denotada  $E(Y | X)$ , no es lineal en  $X$ , el coeficiente de correlación no determinará completamente la forma de  $E(Y | X)$ .

La Figura 2.2 muestra diagramas de dispersión del cuarteto de Anscombe, un conjunto de cuatro pares diferentes de variables creadas por Francis Anscombe. [ Anscombe, Francis J. (1973). "Graphs in statistical analysis". *The American Statistician*. 27: 17–21. JSTOR 2682899. doi:10.2307/2682899] Las cuatro variables y tienen la misma media (7.5), la varianza (4.12), la correlación (0.816) y la línea de regresión ( $y = 3 + 0.5x$ ). Sin embargo, como se puede ver en las imágenes, la distribución de las variables es muy diferente. La primera (superior izquierda) parece estar distribuida normalmente, y corresponde a lo que uno esperaría al considerar dos variables correlacionadas y siguiendo la presunción de normalidad. La segunda (arriba a la derecha) no se distribuye normalmente; Mientras que una relación obvia entre las dos variables puede ser observada, no es lineal. En este caso, el coeficiente de correlación de Pearson no indica que exista una relación funcional exacta: sólo el grado en que esa relación puede ser aproximada por una relación lineal. En el tercer caso (abajo a la izquierda), la relación lineal es perfecta, excepto por un valor atípico que ejerce influencia suficiente para disminuir

el coeficiente de correlación de 1 a 0,816. Finalmente, el cuarto ejemplo (abajo a la derecha) muestra otro ejemplo cuando un valor atípico es suficiente para producir un alto coeficiente de correlación, aunque la relación entre las dos variables no es lineal.

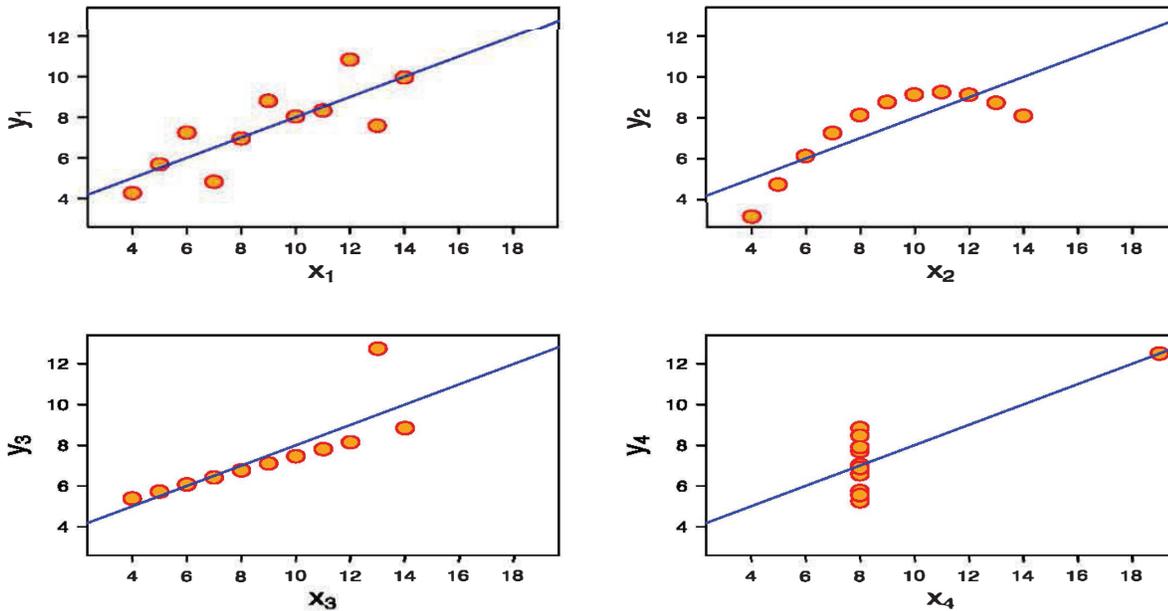


FIGURA 2.2: Dispersión del cuarteto de Anscombe a) Positiva b) Negativa  
c) Sin correlacion d) Correlacion perfecta

Estos ejemplos indican que el coeficiente de correlación, como estadística de resumen, no puede reemplazar el examen visual de los datos. Obsérvese que a veces se dice que los ejemplos demuestran que la correlación de Pearson supone que los datos siguen una distribución normal, pero esto no es correcto.

## 2.2. Técnicas de regresión

### 2.2.1. Regresión lineal

Las técnicas de regresión lineal permiten modelar la relación entre dos variables ajustando una ecuación lineal a los datos observados. Estas técnicas modelan la relación existente entre una variable dependiente y una o más variables explicativas o variables independientes. Cuando se presenta el caso de una sola variable independiente se le conoce como regresión lineal simple. Cuando se presenta el caso de más variables

independientes, el proceso es conocido como regresión lineal múltiple. Existen dos tipos de variables, explicativas o independientes y dependientes, en pocas palabras el modelo desea lograr que las independientes expliquen o modelen el comportamiento medio de la variable independiente, lo veremos con mas detalle a continuacion.

El modelo de Regresion es un modelo estadístico que usa de un conjunto de variables para realizar una ecuación que genere una linea de tendencia que se ajusta a los datos de la variable a explicar. El conjunto de datos de un modelo de regresión se compone de datos de dos o mas variables para que se pueda llevar a cabo el análisis, una variable será la dependiente (VD) mientras que el resto seran variables independientes (VI) y serán las que se ocupen para modelar la variable dependiente; Se debe medir la relación entre ellas no precisamente de causalidad, es decir, que no necesariamente las variables independientes sean la causa del resultado en la variable dependiente y la dependiente sea el efecto de las variables dependientes, sino mas bien se avoca a medir la relacion entre ellas, como medir el comportamiento de una mientras la otra crece o decrece. Puede existir una sola variable independiente, que se ocupará para medir su relación con al VD y diseñar una ecuacion que describa su relacion, al final dicha ecuación puede ser utilizada para predecir el dato promedio siguiente que tendra la VD, o por lo menos un intervalo en el que pueda presentarse el valor futuro de la VD. Tambien se pueden usar mas de una VI y que de manera conjunta se ocuparán para describir a la VD, tantas VI pueden usarse como se desee o necesite, sin embargo uno de los objetivos de este modelo es el desarrollo de una función sencilla que describa en mayor porcentaje la VD con el menor número de VI posible. La funcion se diseñará para poder modelar y predecir los datos de la VD basandose en las entradas que seran las VI.

Si se tiene un ecuación de regresión muy compleja será mas dificil su manejo. Uno de los propositos del modelo de regresión sera generar una ecuacion lo mas sencilla posible, para esto se buscara ocupar modelos ya sean lineales o nó que nos proporcionen un buen nivel de confianza, una buena estimacion y modelado del comportamiento, sin embargo aunque se desea que el modelo sea lo mas preciso posible en sus resultados, tambien se desea que sea un modelo facil de manejar. Por poner un ejemplo si tuvieramos un modelo que variables de entrada  $X_1, X_2, \dots, X_{100}$  que nos describieran la variable dependiente  $Y$  con una precision del 95 %, pero se pudiera generar un modelo con variables  $X_1, X_2, \dots, X_9$  que nos describieran  $Y$  con un 92 % de precision nos inclinariamos por el segundo modelo, ya que el número de variables se esta reduciendo a menos del 10 % del total y el costo en la precision es relativamente pequeño, haciendo

esto tendríamos un modelo mucho más sencillo que unicamnete manejaría 9 variables en su ecuación y no una ecuación que tuviera 100 variables que calcular.

Existen estimadores y parámetros que nos dirán la precisión y nivel de confianza que cada ecuación de regresión generada nos daría, así podremos saber que modelo generado es mejor y que precio estamos dispuestos a pagar por una mejor precisión, generalmente se debe buscar un punto intermedio, ya que tomando el ejemplo anterior podría ser el caso de que únicamente tengamos  $X_i$  y la precisión descienda hasta 80% de nosotros dependerán esas decisiones, aunque por lo general se buscará tener un modelo sencillo con un nivel explicativo alto. La ecuación será lineal en este caso, lo que significa que no posee funciones euler ni exp, ni multiplicaciones o divisiones de alguna variable  $x_i$  entre o por  $x_j$ . Los modelos de regresión puede ser muchísimo más complejos y no necesariamente serán lineales, sino también podrán existir otros muchos modelos de ecuaciones que se puedan ocupar, según el comportamiento de los datos habra funciones que se adecuen mejor a los mismos, si su relación mantiene una pendiente constante y sin curvaturas será una función lineal, si fuera un comportamiento curvo tal vez se requerirá una función cuadrática o de orden superior, según sea el caso ya que aunque nosotros decidimos el orden de la función a ocupar, si elegimos mal, la predicción y/o descripción será sumamente mala.

Formalmente, dado un conjunto de datos  $X = x^1, x^2, \dots, x^n$ , donde  $x^i \in R^d$  y sus etiquetas  $Y = y^1, y^2, \dots, y^n$  donde  $y^i \in R$ . El trabajo del algoritmo de aprendizaje es encontrar una función de salida (hipotesis)  $h_\theta(x)$  que represente lo más cerca posible a la salida  $y$ .

Un modelo de regresión lineal asume que la relación entre la variable dependiente  $y_i$  y el vector  $x_i$  de dimensión  $p$  es lineal. Esta relación es modelada mediante una variable de error  $\epsilon_i$ . Esta variable agrega ruido a la relación lineal entre la variable dependiente y regresores, tomando el modelo la siguiente forma:

$$h_\theta(x) = \theta_0 + \theta_1 x_i \quad (2.3)$$

donde  $\theta_i$  representan los parámetros que configuran el espacio de funciones lineales mapeando de  $X$  a  $Y$ , estos parámetros también son llamados pesos,  $\theta_1$  representa la pendiente de la línea,  $\theta_0$  el sesgo o intercepción de la línea (el valor de  $y$  cuando  $x = 0$ ) y  $x$  es la variable explicativa.

Cuando desconocemos los pesos que definen a nuestra hipótesis  $h_\theta(x)$  y necesitamos conocerlos. Un método razonable es reducir el error entre la hipótesis y la salida real  $y$  (también llamado objetivo). Para formalizar esto vamos a definir una función que

mida, para cada valor de  $\theta$ , el error obtenido de la diferencia entre las hipótesis  $h_\theta(x_i)$ s con sus correspondientes objetivos reales  $y^{(i)}$ 's.

$$\min_{\theta_0\theta_1} \frac{1}{2n} \sum_{i=1}^n (h_\theta(x^{(i)}) - y^{(i)})^2 \quad (2.4)$$

Este método es utilizado cuando queremos resolver un problema de regresión simple (una sola variable independiente). La idea de la regresión es elegir  $\theta_0$  y  $\theta_1$  de tal forma que  $h_\theta(x)$  sea lo más cercana a la salida real  $y$  definida en nuestro conjunto de datos. En otras palabras, tenemos que minimizar la ecuación 2.4. Esta ecuación la conocemos como función de costo y es una función de costo de mínimos cuadrados.

$$J(\theta_0\theta_1) = \frac{1}{2n} \sum_{i=1}^n (h_\theta(x^{(i)}) - y^{(i)})^2 \quad (2.5)$$

El objetivo en regresión lineal simple es minimizar esta función de costo  $\min_{\theta_0\theta_1} J(\theta_0\theta_1)$ . Existen otras funciones de costo que trabajan muy bien al resolver problemas de regresión, pero la función de costo de error cuadrático es probablemente la más utilizada para resolver problemas de regresión.

### 2.2.2. Algoritmo Least-Mean-Square

El objetivo del algoritmo es encontrar un conjunto de pesos  $\theta$  que minimicen la función de costo  $J(\theta)$ . Para llevar a cabo esto, utilizaremos un algoritmo de búsqueda que iniciará con algunos pesos aleatorios para  $\theta$  y que de forma iterativa cambiará  $\theta$  para hacer  $J(\theta)$  más pequeño, hasta que el algoritmo converja a un valor de  $\theta$  que minimice  $J(\theta)$ . Para ello, utilizaremos el algoritmo de gradiente descendiente, que inicia con un conjunto de valores de  $\theta$  iniciales e iterativamente actualiza su valor mediante:

$$\theta_0 := \theta_0 - \alpha \frac{\partial}{\partial \theta_0} J(\theta_0\theta_1) \quad (2.6)$$

Esta actualización es realizada simultáneamente para  $\theta_0$  y  $\theta_1$ . En la ecuación 2.6,  $\alpha$  es el parámetro de aprendizaje. El algoritmo permite iterativamente converger a la solución en la dirección del gradiente decreciente de  $J(\theta)$ .

Con el objetivo de implementar este algoritmo, es necesario obtener la derivada parcial de la función de costo. Consideremos el caso de un solo ejemplo en entrenamiento  $(x, y)$ , de tal forma que no tengamos que hacer la sumatoria de las diferencias de la función de costo.

$$\frac{\partial}{\partial \theta_j} J(\theta) = \frac{\partial}{\partial \theta_j} \frac{1}{2} (h_\theta(x) - y)^2 \quad (2.7)$$

$$= 2 \cdot \frac{1}{2} (h_\theta(x) - y) \cdot \frac{\partial}{\partial \theta_j} (h_\theta(x) - y) \quad (2.8)$$

$$= (h_\theta(x) - y) \cdot \frac{\partial}{\partial \theta_j} \left( \sum_{i=0}^n \theta_i x_i - y \right) \quad (2.9)$$

$$= (h_\theta(x) - y) x_j \quad (2.10)$$

Para solo un ejemplo de entrenamiento, los pesos son actualizados de la siguiente forma:

$$\theta_j := \theta_j + \alpha (y^{(i)} - h_\theta(x^{(i)})) x_j^{(i)} \quad (2.11)$$

Este algoritmo de actualización es llamado Least-Mean-Squares (LMS) y también es conocido como algoritmo de aprendizaje de Widrow-Hoff. Este algoritmo posee varias propiedades naturales e intuitivas. La magnitud de la actualización es proporcional al error  $(y^{(i)} - h_\theta(x^{(i)}))$ . Si el valor obtenido por nuestra hipótesis  $h_\theta(x^{(i)})$  es muy cercano a nuestra salida real  $y^{(i)}$ , entonces los parámetros deben ser modificados muy poco; sin embargo, si el valor de nuestra hipótesis es muy diferente a nuestra salida real, el error será muy grande y los parámetros deberán ser modificados para disminuir la diferencia entre la predicción  $h_\theta(x^{(i)})$  y la salida real  $y^{(i)}$ .

Sin embargo, las ecuaciones anteriores permiten obtener la hipótesis cuando tenemos una sola característica independiente. Sin embargo, la mayoría de los problemas del mundo real tienen múltiples características de entrada, es decir la hipótesis es definida mediante:

$$h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \dots + \theta_n x_n \quad (2.12)$$

Para simplificar la notación, asignamos  $x_0 = 1$ , de tal forma que

$$h_\theta(x) = \sum_{i=0}^n \theta_i x_i = \theta^T \mathbf{x} \quad (2.13)$$

donde  $^T$  denota la transpuesta, de tal forma que  $\mathbf{x}_i^T \theta$  es el producto interno entre vectores  $\mathbf{x}_i$  y  $\theta$ .

$$\text{donde } \mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \mathbf{x} = \begin{pmatrix} \mathbf{x}_1^T \\ \mathbf{x}_2^T \\ \vdots \\ \mathbf{x}_n^T \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & \dots & x_{1p} \\ 1 & x_{21} & \dots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \dots & x_{np} \end{pmatrix}, \boldsymbol{\theta} = \begin{pmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \\ \vdots \\ \theta_p \end{pmatrix}$$

El algoritmo es iterativamente actualizado para converger a la solución utilizando:

$$\theta_j := \theta_j - \alpha \sum_{i=0}^m (y^{(i)} - h_{\theta}(x^{(i)})) x_j^{(i)} \quad (2.14)$$

donde  $m = n + 1$ . El algoritmo 1 describe el algoritmo de regresión para múltiples variables independientes.

---

**Algorithm 1** Algoritmo de regresión lineal simple (Múltiples variables independientes)

---

**Require:** Parámetros iniciales  $\theta_0, \theta_1, \dots, \theta_j$  y  $\alpha$

**Ensure:** Parámetros aprendidos  $\theta_0, \theta_1, \dots, \theta_j$

1: **while** Los parámetros  $\theta_j$  no converjan **do**

2:  $\theta_j := \theta_j - \alpha \sum_{i=0}^m (y^{(i)} - h_{\theta}(x^{(i)})) x_j^{(i)}$

3: **end while**

---

## 2.3. Neural networks

Las redes neuronales artificiales (RNA) han recibido gran atención en los últimos años. Estas se han implementado para resolver problemas de predicción y clasificación, áreas en las que tradicionalmente se han utilizado técnicas estadísticas. Sin embargo, en los últimos años han obtenido una inmensa popularidad para resolver problemas de regresión.

Las RNA construyen un modelo de predicción imitando la inteligencia del cerebro humano. De forma similar a como lo hace el cerebro, las RNA identifican regularidades y patrones en los datos, aprenden a partir de la experiencia y proveen resultados que son generalizados a partir del conocimiento obtenido. Una red neuronal es un conjunto de neuronas artificiales interconectadas que utilizan modelos matemáticos para procesar información. Las múltiples conexiones entre las neuronas forman un sistema adaptivo cuyos pesos se actualizan mediante un particular algoritmo de aprendizaje. Las redes neuronales han sido utilizadas en numerosos campos de aplicación con distintos algoritmos de aprendizaje **Portillo:2009 Jimenez:2013 G:2010 Valverde:2007**

### 2.3.1. Elementos de las RNA

En los últimos años, debido al auge de las RNA se han utilizado diferentes RNA para resolver problemas de predicción. Las características esenciales de una RNA son las siguientes:

1. Los elementos básicos de procesamiento (neuronas o nodos)
2. La arquitectura de la red describiendo las conexiones entre nodos
3. El algoritmo de entrenamiento usado para encontrar valores de los parámetros de la red

Una RNA consiste de elementos de procesamiento básico (neuronas) organizadas en capas. Las capas que se encuentran entre la capa de entrada y capa de salida son llamadas capas ocultas. El número de neuronas en la capa de entrada es determinada por la aplicación. La arquitectura o topología de una RNA se refiere al arreglo de las conexiones de la red. Una RNA es especificada mediante:

$$S = \{h(x, \theta), x \in R^m, \theta_i \in \theta\}, \theta \subseteq R^p \quad (2.15)$$

donde  $h_\theta(x, \theta)$  es una función no lineal del producto punto de  $x$  con  $\theta$ , es el número de neuronas ocultas,  $\theta = (\theta_1, \theta_2, \dots, \theta_p)^T$  es un vector cuyos elementos usualmente son llamados pesos y  $p$  es el número de parámetros libres.

### 2.3.2. Función de costo y algoritmo de aprendizaje

Dado un conjunto de observaciones, la tarea de una RNA consiste en construir un estimador  $g(x, \theta)$  de la función desconocida  $\varphi(x)$

$$g_\lambda(x, \theta) = f_2 \left( \sum_{j=1}^{\lambda} \theta_j^2 f_1 \left( \sum_{i=1}^m \theta_{ij}^1 x_i + \theta_{m+1,j}^1 \right) + \theta_{\lambda+1}^1 \right) \quad (2.16)$$

donde  $\theta = (\theta_1, \theta_2, \dots, \theta_p)^T$  es el vector de parámetros a ser estimados,  $f_i$  es conocida como la función de activación, Las funciones de activación comunmente utilizadas en RNA son función sigmoideal o tangente hiperbolica.

Los parámetros estimados  $\hat{\theta}$  son obtenidos minimizando iterativamente una función de costo

$$J(\theta) = \frac{1}{2n} \sum_{j=1}^n (y_j - g_\lambda(x_j, \theta))^2 \quad (2.17)$$

El requerimiento básico de cualquier método es la convergencia del algoritmo de entrenamiento a un mínimo local. La función de costo nos da una medida de la precisión con la que el estimador  $g_\lambda(x_i, \theta)$  ajusta los datos observados. Para llevar a cabo el proceso de aprendizaje, el algoritmo cambia iterativamente los pesos entre las neuronas minimizando el error cuadrático entre la salida deseada y la obtenida con los pesos actuales. Cada uno de los ejemplos del conjunto de entrenamiento  $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$  son utilizados para ajustar los pesos en la red. Al ser presentado un ejemplo, la señal es propagada hacia adelante de la red hasta que la salida es obtenida. La salida de la  $j$ -ésima unidad oculta es calculada como:

$$g(x, \theta)_{nj}^h = f_j^h(\text{net}_{nj}^h) = \frac{1}{1 + \exp(-\text{net}_{nj}^h)} \quad (2.18)$$

donde  $\text{net}_{nj}^h = \sum \theta_{ji}^h x_{ni}$ ,  $\theta_{j0}^h$  es el sesgo cuyo  $x_0 = 1$ .  $\theta_{ji}^h$  es el peso de la conexión de la  $i$ -ésima neurona de entrada a la  $j$ -ésima neurona oculta.  $f_j^h$  representa la función de activación de la  $j$ -ésima neurona oculta. Por otro lado, la salida de la  $k$ -ésima neurona está dado por

$$g(x, \theta)_{nk}^o = f_k^o(\text{net}_{nk}^o) = \frac{1}{1 + \exp(-\text{net}_{nk}^o)} \quad (2.19)$$

donde los superíndices  $h$  y  $o$  se refieren a las cantidades en las capas ocultas y de salida respectivamente. El error entre la salida actual y la salida deseada es calculado para ajustar los pesos mediante  $E_n = \frac{1}{2} \sum_{k=1}^C (y_{nk} - g(x, \theta)_{nk}^o)$ . El procedimiento de ajuste es obtenido a partir del método de gradiente descendente para reducir la magnitud del error. El procedimiento es primeramente aplicado a los pesos en la capa de salida y retropropagado a través de la red hasta que los pesos en la primera capa han sido ajustados  $\Delta\theta_{kj}^o = -\eta \frac{\partial E_n}{\partial \theta_{kj}^o}$  y  $\Delta\theta_{ji}^h = -\eta \frac{\partial E_n}{\partial \theta_{ji}^h}$ . Este procedimiento es realizado para cada ejemplo en el conjunto de datos hasta que se cumple un criterio de paro. Para un estudio profundo del algoritmo puede referirse a **Rumel:1986 Werbos:1994**

### 2.3.3. Modelo de RNA para predicción

Las RNA pueden ser clasificadas como dispositivos para inferencia estadística no-paramétricos. Dado un conjunto de ejemplos  $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$  donde  $x_1 \in R^d$  y  $y_i \in R$ , generados por una función desconocida  $f(x)$  con la adición de un componente estocástico  $\epsilon$ , *esto es*

$$y = f(x) + \epsilon_i \quad (2.20)$$

La tarea del aprendizaje neuronal es construir un estimador  $g(x, \theta) \equiv \wedge f(x)$  de  $f(x)$ , donde  $x_i$  es un conjunto finito de observaciones. Ya que no conocemos el comportamiento a priori de la función  $f(x)$ , el modelo neuronal  $g(x, \theta)$  es un estimador no paramétrico.

## 2.4. Support Vector Machines

Las Maquinas de vectores de Soporte (SVM por sus siglas en ingles) son una de las técnicas de clasificación y regresión más utilizadas en los últimos años. Las características clave de las SVMs son el uso de *kernels* al trabajar en conjuntos no-lineales, la ausencia de los mínimos locales, aunado a ello la solución depende de un pequeño subconjunto de datos y el poder discriminativo del modelo obtenido es significativamente mayor a otras técnicas al optimizar el margen de separabilidad entre clases, estas características permiten a las SVM obtener resultados muy competitivos en comparación con otros clasificadores.

Las SVM fueron desarrolladas por Vapnik y sus colegas en los laboratorios de ATT en 1995 **Vapnik:1998** En un inicio las SVM fueron diseñadas para resolver problemas de clasificación como reconocimiento de caracteres y reconocimiento facial. Sin embargo, pronto se fueron utilizando no solo para resolver problemas de clasificación sino también para resolver problemas de regresión en variadas implementaciones y múltiples dominios **Vapnik:1998**

Las SVM son basadas en el principio de minimización del riesgo estructural. El objetivo de las SVM es encontrar una hipótesis  $h_\theta(x)$  con una buena habilidad de generalización a partir de un conjunto de entrenamiento. Esta hipótesis es completamente definida a partir de un pequeño subconjunto del conjunto de entrenamiento original. En las siguientes subsecciones se describe formalmente a las SVM.

### 2.4.1. Teoría de aprendizaje estadístico

La teoría de aprendizaje estadístico es desarrollada con el objetivo de obtener una técnica de aprendizaje con una buena capacidad de generalización. En problemas de clasificación y regresión el objetivo es encontrar una hipótesis (función) a partir de los

datos de entrenamiento y usando una máquina de aprendizaje que infiera resultados basados en este conocimiento.

En el caso de aprendizaje supervisado, los datos de entrenamiento son compuestos por pares de entrada y salida. Los vectores de entrada  $x \in X \subseteq R^n$  y los puntos de salida  $y \in Y \subseteq R$ . Los dos subconjuntos  $X$  e  $Y$  son definidos como espacios de entrada y salida respectivamente.  $Y = -1, 1$  o  $0, 1$  para problemas de clasificación binaria y  $Y = R$  para problemas de regresión.

Es claro que los datos de entrenamiento son generados a partir de una distribución desconocida  $P(\mathbf{x}, y)$  definida sobre un conjunto  $X \times Y$ . Un vector de entrada es trazado desde  $X$  con la probabilidad marginal  $P(x)$  y su correspondiente punto de salida en  $Y$  con la probabilidad condicional  $P(\mathbf{x}, y)$ .

Después de estas descripciones, el problema de aprendizaje puede ser visualizado como la búsqueda de una apropiada función de aproximación  $f : X \rightarrow Y$  que represente el proceso de obtención de salidas a partir de los vectores de entrada. Esta función puede ser utilizada para generalizar, es decir para producir una salida a partir de un vector de entrada nunca antes visto por el modelo.

### 2.4.2. Minimización del riesgo empírico

Según Vapnik **Vapnik:1998** el riesgo funcional es definido sobre  $X \times Y$  para medir el error promedio obtenido de las salidas real y predecida al utilizar una función de aproximación  $f$ . La función de aproximación mas adecuada es seleccionada como la función que minimiza este riesgo.

Considerando un conjunto de funciones  $F = f(\mathbf{x}, \mathbf{w})$  que mapea los puntos de un espacio de entrada  $X \subseteq R^n$  dentro de un espacio de salida  $Y \subseteq R$  donde  $w$  denota los parámetros que definen a  $f$ .

Suponiendo que  $y$  sea el punto actual de salida correspondiente al vector de entrada  $\mathbf{x}$ . Ahora si  $L(y, f(\mathbf{x}, \mathbf{w}))$  mide el error entre el valor actual  $y$  y el valor predecido  $f(\mathbf{x}, \mathbf{w})$  usando la función de predicción  $f$  entonces el riesgo esperado es definido como:

$$R(f) = \int L(y, f(\mathbf{x}, \mathbf{w})) dP(\mathbf{x}, y) \quad (2.21)$$

donde  $P(\mathbf{x}, y)$  es la distribución de probabilidad de los datos de entrenamiento.  $L(y, f(\mathbf{x}, \mathbf{w}))$  es conocido como la función de perdida y esta puede ser definida como un número de soluciones.

La función de predicción mas adecuada es aquella que minimiza el riesgo esperado  $R(f)$  y es denotada como  $f_0$ . Esta es conocida como la función objetivo. La principal tarea del problema de aprendizaje es ahora encontrar la función objetivo, que es el estimador ideal. Desafortunadamente esto no es posible debido a que probabilidad de distribución  $P(\mathbf{x}, y)$  de los datos dados es desconocida y por lo tanto el riesgo esperado no puede ser calculado. Este problema motivo a Vapnik a sugerir el principio de minimización de riesgo empírico (MRE) [1]. El concepto de MRE estima el riesgo esperado  $R(f)$  usando el conjunto de entrenamiento. Esta aproximación de  $R(f)$  es llamada el riesgo empírico. Dado un conjunto de entrenamiento  $(\mathbf{x}_i, y_i)$ , donde  $\mathbf{x}_i \in X \subseteq R^n$ ,  $y_i \in Y \subseteq R$  ( $\forall i = 1, 2, \dots, n$ ) el riesgo empírico es definido como:

$$R_{emp}(f) = \frac{1}{n} \sum_{i=1}^n L(y_i, f(\mathbf{x}_i, \mathbf{w})) \quad (2.22)$$

El riesgo empírico  $R_{emp}(f)$  tiene su propio minimizador en  $F$ , que puede ser tomado como  $\hat{f}$ . El objetivo de MRE es aproximar la función objetivo  $f_0$  mediante  $\hat{f}$ . esto es posible debido a que  $R(f)$  converge en  $R_{emp}(f)$  cuando el conjunto de entrenamiento  $n$  es infinitamente grande.

### 2.4.3. Dimensión Vapnik-Chervonenkis (VC)

La dimensión Vapnik-Chervonenkis  $h$  de una clase de funciones  $F$  es definida como el máximo número de puntos que pueden ser exactamente clasificados mediante  $F$ . Matemáticamente:  $h = \max |X|$ ,  $X \subseteq R^n$  tal que  $\forall b \in \{-1, 1\}^{|X|}$ ,  $\exists f \in F$  tal que  $\forall x_i \in X (1 \leq i), f(x_i) = b_i$ .

La dimensión VC es una medida de la capacidad intrínseca de una clase de funciones  $F$ . Burgues [1998] menciona que la dimensión VC de un conjunto de hiperplanos orientados en  $R^n$  es  $(n + 1)$ . Esto es, tres puntos etiquetados en 8 diferentes formas pueden ser siempre clasificados por una frontera de decisión orientada en  $R^2$  pero cuatro puntos no. Esto es la dimensión VC del conjunto de líneas orientadas en  $R^2$  es tres. Por ejemplo el problema XOR no puede ser resuelto utilizando frontera de decisión lineal. Sin embargo, una frontera de decisión cuadrática puede correctamente clasificar los puntos en este problema.

### 2.4.4. Minimización del Riesgo Estructural MRE

El defecto principal del principio de MRE es que en la practica siempre tenemos un conjunto finito de observaciones y no puede ser garantizado que minimizando la función de aproximación del riesgo empírico sobre  $F$  tambien minimizaremos el riesgo esperado. Para enfrentar esta desventaja fue desarrollado el principio de minimización de riesgo estructural por Vapnik y Chervonenkis en 1982. La clave de este principio es que la diferencia entre el riesgo empirico y riesgo esperado puede ser acotada en terminos de de la dimensión VC de la clase  $F$  de las funciones de aproximación.

**Teorema:** Sea  $F$  una clase de funciones de aproximación de la dimensión  $h$ . Entonces para cualquier par  $(\mathbf{x}_i, y_i)$ , donde  $\mathbf{x}_i \in X \subseteq R^n, y_i \in Y \subseteq R (\forall i = 1, 2, \dots, n)$  trazadas para cualquier distribución  $P(\mathbf{x}, y)$  la siguiente cota se mantiene con probabilidad  $1 - \eta (0 \leq \eta \leq 1)$  :

$$R(f) \leq R_{emp}(f) + \sqrt{\frac{h \left( \ln \left( \frac{2n}{h} \right) + 1 \right) - \ln \left( \frac{n}{4} \right)}{n}} \quad (2.23)$$

El segundo término del lado derecho se dice que es la confianza VC y  $1 - \eta$  se denomina el nivel de confianza. El principio de MRE sugiere que para llevar a cabo una buena generalización es necesario minimizar la combinación del riesgo empírico y la complejidad del espacio de hipótesis. En otras palabras, es necesario seleccionar una hipótesis que tenga una buena relación entre un pequeño error empírico y una pequeña complejidad del modelo.

### 2.4.5. Support Vector Machines

La principal idea de SVM en clasificación binaria es encontrar un hiperplano canónico que maximice la separación entre dos clases de ejemplos de entrenamiento. Consideremos dos conjuntos de puntos que son linealmente separables en  $R^n$  los cuales son clasificados dentro de una de dos clases  $C_1$  y  $C_2$  utilizando hiperplanos lineales. A partir del conjunto infinito de hiperplanos de separación es seleccionado aquel con el máximo margen de separación, que es el que tiene la mejor capacidad de generalización.

Para una mejor precisión así como una mejor capacidad de generalización, el hiperplano que maximiza el margen total es considerado como el óptimo y conocido como el Hiperplano de Margen Máximo. Es claro que para este hiperplano óptimo  $d_+ = d_-$ .

Los datos ya sea de una u otra clase que se encuentran más cercanos al hiperplano son conocidos como vectores de soporte.

Considerando que el conjunto de entrenamiento es compuesto por los pares de entrada salida

$$(\mathbf{x}_1, \mathbf{y}_1), (\mathbf{x}_2, \mathbf{y}_2), \dots, (\mathbf{x}_n, \mathbf{y}_n) \quad (2.24)$$

i.e.  $X = \{x_i, y_i\}_{i=1}^n$  donde  $x_i \in R^d$  y  $y_i \in (+1, -1)$ .

El objetivo es clasificar los datos en dos clases encontrando el hiperplano canonico de máximo margen. el espacio de hipotesis es el conjunto de funciones  $f(\mathbf{x}, \theta, b) = \text{sgn}(\theta^T \mathbf{x} + b)$  donde  $\theta$  es el vector de pesos,  $x \in R^n$  y  $b$  es el sesgo. el conjunto de hiperplanos de separación es dado por  $\{x \in R^n : (\theta^T \mathbf{x} + b) = 0\}$ , donde  $\theta \in R^n, b \in R$ . Usando SVM para encontrar el hiperplano de separación máximo se acota el problema a resolver un problema de programación cuadrática (QPP).

#### 2.4.6. SVM para datos linealmente separables

Para datos linealmente separables el problema de optimización cuadrática es dado por

$$\begin{aligned} \text{mín } J(\theta) &= \frac{1}{2} \theta^T \theta = \frac{1}{2} \|\theta\|^2 \\ \text{subject to } &y_i(\theta^T \mathbf{x}_i + b); \forall i = 1, 2, \dots, n \end{aligned}$$

Para resolver el problema de optimización cuadrática es necesario transformarlo al espacio dual. Entonces los multiplicadores de Lagrange y las condiciones complementarias de Kuhn-Tucker son usadas para encontrar la solución óptima.

Considerando que la solución del problema de Optimización Cuadrática nos genera los multiplicadores de Lagrange optimizados  $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_n)^T$ , donde  $\alpha_i \geq 0$  y el sesgo óptimo esta dado por  $b_{opt}$ . Los vectores de datos para los que  $\alpha_i > 0$  son los vectores soporte y se supone que existen en total  $P_{sv}$  vectores soporte. Entonces el vector de pesos óptimo puede ser escrito como:

$$\theta_{opt} = \sum_{i=1}^{P_{sv}} \alpha_i y_i \mathbf{x}_i \quad (2.25)$$

La función del hiperplano de decisión óptimo es dado como:

$$y(\mathbf{x}) = \text{sgn} \left( \sum_{i=1}^{P_{sv}} y_i \alpha_i (\mathbf{x}^T \mathbf{x}_i) + b_{opt} \right) \quad (2.26)$$

Un dato desconocido es clasificado en una de las dos clases de acuerdo a la ecuación (2.26).

### 2.4.7. SVM para datos linealmente no separables

En aplicaciones de la vida real los datos de entrenamiento son linealmente no-separables. Como ejemplo podemos tomar el problema de clasificación XOR. En tales casos un Hiperplano de Margen Suave es modelado. El problema de optimización cuadrática en tal caso es dado por

$$\begin{aligned} \text{mín } J(\theta) &= \frac{1}{2} \|\theta\|^2 + C \sum_{i=1}^n \xi_i \\ \text{subject to } & y_i (\theta^T \mathbf{x}_i + b) \geq 1 - \xi_i; \forall i = 1, 2, \dots, n \end{aligned}$$

Aquí las variables flojas  $\xi_i$  son introducidas para relajar las condiciones de margen duro y  $C > 0$  es la constante de regularización que asigna una penalización a las clasificaciones erróneas. El vector de pesos óptimo y la función de decisión es similar al caso de separabilidad lineal. La única diferencia es que en este caso los multiplicadores de Lagrange son acotados por  $C$  i.e.  $0 \leq \alpha_i \leq C$  y por los vectores soporte.

### 2.4.8. Kernels

Las SVM permiten mapear el espacio de los puntos de entrada a un espacio de características altamente dimensional a través de un mapeo no-lineal. El hiperplano de separación no-lineal es construido en este nuevo espacio de características. Este truco permite resolver el problema de los datos de entrenamiento cuando estos no son separables linealmente. Al utilizar una transformación apropiada los datos de entrada pueden ser separables linealmente en el espacio de características.

Debido a esta transformación, los datos linealmente no-separables pueden ser separados por una frontera de decisión  $\pi_f$  en el espacio de características. La frontera de decisión en  $H$  puede ser reescrita como:

$$y(\mathbf{x}) = \text{sgn} \left( \sum_{i=1}^{P_{sv}} y_i \alpha_i (\varphi(\mathbf{x})^T \varphi(\mathbf{x}_i)) + b_{opt} \right) \quad (2.27)$$

Si es posible encontrar una función  $\mathbf{K}(x_i, x_j) = \varphi(x_i^T)\varphi(x_j)$  entonces esta puede ser directamente utilizada en las ecuaciones de entrenamiento de una SVM, esta función es conocida como la función kernel. Para evitar el computo explícito de del mapeo no-lineal  $\varphi(x)$ , el kernel asociado debe satisfacer las condiciones de Mercer **Vapnik:1998** Los kerneles más conocidos que han sido utilizados en las SVM son los siguientes:

1. Kernel lineal  $\mathbf{K}(\mathbf{x}, \mathbf{y}) = x^T y$
2. Kernel polinomial  $\mathbf{K}(\mathbf{x}, \mathbf{y}) = (1 + x^T y)^d$
3. Kernel Función radial base (RBF)  $\mathbf{K}(\mathbf{x}, \mathbf{y}) = \exp\left(\frac{-\|x-y\|^2}{2\sigma^2}\right)$  que en forma simple puede ser escrito como  $\exp\left(-\gamma \cdot \|X^T Y\|^2\right)$ ,  $\gamma > 0$ .
4. Kernel sigmoidal  $\mathbf{K}(\mathbf{x}, \mathbf{y}) = \tanh(\gamma \cdot X^T Y + r)$  que es similar a la función sigmoidal en regresión logística.

### 2.4.9. SVM para regresión (SVR)

En SVR, Vapnik utilizó la función de perdida  $\xi$  - sensitive definida como:

$$L_\xi(y, f(\mathbf{x}, \theta)) = \begin{cases} 0 & \text{if } |y - f(\mathbf{x}, \theta)| \leq \epsilon \\ |y - f(\mathbf{x}, \theta)| - \epsilon & \text{de otra manera} \end{cases} \quad (2.28)$$

Entonces el riesgo empírico puede ser minimizado mediante

$$R_{emp}(f) = \frac{1}{n} \sum_{i=1}^n L_\xi(y_i, f(\mathbf{x}_i, \theta)) \quad (2.29)$$

El problema de programación cuadrática asociado puede ser escrito como:

$$\begin{aligned} \text{mín } J(\theta, \xi, \xi^*) &= \frac{1}{2} \|\theta\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*) \\ \text{subject to} & \\ y_i(\theta^T \varphi(\mathbf{x}_i) - b) &\leq \epsilon + \xi_i; \forall i = 1, 2, \dots, n \\ \theta^T \varphi(\mathbf{x}_i) + b - y_i &\leq \epsilon + \xi_i^* \\ \xi_i &\geq 0 \\ \xi_i^* &\geq 0 \end{aligned} \quad (2.30)$$

Para obtener la solución de 2.30 dos conjuntos de multiplicadores de Lagrange son utilizados, estos son  $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_n)^T$  y  $\alpha^* = (\alpha_1^*, \alpha_2^*, \dots, \alpha_n^*)^T$  donde  $0 \leq \alpha_i, \alpha_i^* \leq C$ . Para los vectores soporte se cumple que  $\alpha_i > 0, \alpha_i^* \leq C$ . Finalmente, el hiperplano de decisión óptima es obtenido como:

$$y(\mathbf{x}) = \sum_{i=1}^n (\alpha_i - \alpha_i^*) k(\mathbf{x}, \mathbf{x}_i) + b_{opt} \quad (2.31)$$

Durante los últimos años se han llevado a cabo muchos trabajos de investigación sobre predicción de series de tiempo utilizando SVM, esto ha provocado que existan en la literatura excelentes algoritmos de predicción utilizando SVM.

### 2.4.10. Least Square Support Vector Machine (LS-SVM)

Dado un conjunto de entrenamiento de  $n$  puntos  $\{\mathbf{x}_i, y_i\}_{i=1}^n$  con datos de entrada  $\mathbf{x}_i \in R^n$  y la respuesta  $y_i \in R$ . Entonces el problema de optimización esta dado por:

$$\begin{aligned} \text{mín } \theta, b, e, J(\theta, e) &= \frac{1}{2} \theta^T \theta + \frac{1}{2} \gamma \sum_{i=1}^n e_i^2 \\ \text{subject to } y_i &= \theta^T \varphi(\mathbf{x}_i) + b + e_i; \forall i = 1, 2, \dots, n \end{aligned} \quad (2.32)$$

donde  $\varphi$  es el mapeo no lineal a un espacio altamente dimensional y  $\gamma$  es el parámetro de regularización. El kernel RBF  $\mathbf{K}(\mathbf{x}, \mathbf{y}) = \exp\left(\frac{-\|\mathbf{x}-\mathbf{y}\|^2}{2\sigma^2}\right)$  que en forma simple puede ser escrito como  $\exp\left(-\gamma \cdot \|\mathbf{X}^T \mathbf{Y}\|^2\right)$ ,  $\gamma > 0$  con el parámetro  $\gamma$  puede ser empleado. El problema de optimización en el espacio primal esta dado por:

$$y = \theta^T \varphi(\mathbf{x}) + b \quad (2.33)$$

Por simplicidad computacional y para evitar el caso de dimensionalidad infinita del vector de pesos  $\theta$  las operaciones de optimización son realizadas en el espacio dual. El Lagrangiano del problema esta dado por:

$$L(\theta, b, e; \alpha) = J(\theta, e) - \sum_{i=1}^n \alpha_i \left\{ \theta^T \varphi(\mathbf{x}_i) + b + e_i - y_i \right\} \quad (2.34)$$

donde  $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_n)^T$ ,  $\alpha_i > 0 (\forall i = 1, 2, \dots, n)$  son los multiplicadores de Lagrange.

Aplicando las condiciones de optimalidad y calculando las derivadas parciales de  $L$  con respecto a  $\theta, b, e_k; \alpha_k$  e igualando las ecuaciones resultantes con cero obtenemos el siguiente sistema de ecuaciones:

$$\begin{bmatrix} 0 & 1 \\ 1^T & \Omega + \gamma^{-1} I \end{bmatrix}_{(n+1) \times (n+1)} \begin{bmatrix} b \\ \alpha \end{bmatrix}_{(n+1) \times 1} = \begin{bmatrix} o \\ y \end{bmatrix}_{(n+1) \times 1} \quad (2.35)$$

donde  $\mathbf{y} = [y_1, y_2, \dots, y_n]$ ,  $\mathbf{1} = [1, 1, \dots, 1]$  y  $\Omega$  con  $\Omega(i, j) = k(\mathbf{x}_i, \mathbf{x}_j) (\forall i, j = 1, 2, \dots, n)$  es la matriz kernel. La función de decisión es entonces dada como:

$$y(x) = \sum_{i=1}^n \alpha_i k(\mathbf{x}, \mathbf{x}_i) + b \quad (2.36)$$

donde  $\alpha, b$  son las soluciones del sistema lineal 2.4.10. El principal beneficio de la técnica LS-SVM radica en que esta transforma la tradicional QPP a un problema de sistemas lineales, asegurando la simplicidad de los cálculos, rápida convergencia y alta precisión.

## 2.5. Selección de características utilizando un Algoritmo Genético

Las características utilizadas en los conjuntos de datos regularmentes son muchas, esto es debido a que desconocemos que características influyen en la predicción. En nuestro caso, hay variables económicas, sociales, demográficas, actividades agropecuarias e industriales, entre otras muchas. Sin embargo, muchas de estas variables no son importantes, no benefician y en ocasiones en lugar de influir positivamente en la predicción, afectan el desempeño del predictor, debido a que añaden ruido colocando valores que confunden al predictor.

El uso adecuado de características para una buena clasificación ha sido tratado por varios autores y lo llaman, el curso de la dimensionalidad. El curso de la dimensionalidad nos dice que la precisión de un clasificador/predictor va aumentando conforme aumenta el número de características, pero llegado a un cierto número de características, la precisión del predictor/clasificador va decayendo conforme sigue aumentando el número de características. La Figura muestra el ejemplo del curso de dimensionalidad y se ve claramente que al aumentar las características, aumenta la precisión de clasificación hasta cierto punto, ya que después la empeora con mayor impacto. Así que el desafío de reducir la dimensionalidad de características es fundamental en la etapa de clasificación.

Un factor importante al momento de reducir características, es eliminar aquellas que no son clave importante en el predictor. Aunque, el hecho de decir que cierto descriptor aporta ruido en un clasificador, no quiere decir que no funcione o que sea inútil en otras clasificaciones, más bien significa que tal vez es un rasgo o características que comparte el mismo valor con otras características. Si dicho descriptor se utilizará para

identificar otros objetos, tal vez pueda ser un buen elemento y tener buena efectividad en su identificación.

Con el objetivo de identificar las características menos eficientes, se implementó un algoritmo genético básico sobre las características de cada uno de los conjuntos de datos. La forma en que se adaptó el genético (AG) se describe a continuación:

Los algoritmos genéticos (AGs), parten de la premisa de emplear la evolución natural como un procedimiento de optimización. Se caracterizan por representar las soluciones al problema que abordan en forma de cadenas binarias. Esas representaciones binarias les aportan características muy importantes de eficiencia. Sin embargo, es necesario disponer de un método para pasar esa representación binaria al espacio de búsqueda natural de cada problema.

En esta sección, describimos una técnica para eliminar las características menos importantes, que influyan poco o negativamente en la predicción.

### 2.5.1. Elementos de un algoritmo genético

Para ejecutar un AG, se requiere de una población de individuos. Cada individuo, es un candidato a ser la solución del problema tratado, o permite llegar a la solución a partir de este.

Cada individuo de la población se representa con una cadena binaria y se denomina *genotipo* del individuo que es análoga al *cromosoma* en el sistema biológico. Cada genotipo representa a  $x$  puntos del espacio de búsqueda del problema. A cada punto  $x$  se le denomina *fenotipo*. Se usa el término *gen* para referirse a la codificación de una determinada característica del individuo. Cada *gen* puede tomar distintos valores que son llamados *alelos*. Para referirse a una determinada posición de la cadena binaria se usa el término locus. La Tabla 2.1 muestra estas expresiones que se usan comúnmente en la genética y su estructura equivalente en un algoritmo genético:

CUADRO 2.1: Expresiones que se utilizan en la genética con su estructura equivalente en un algoritmo genético

Evolución natural	Algoritmo genético
cromosoma	cadena
genotipo	código de cadena
fenotipo	punto sin codificar
gen	posición de cadena
alelo	valor en una posición determinada
aptitud	valor de la función objetivo

### 2.5.2. Algoritmo genético básico

En algoritmos genéticos, el termino básico o simple, es debido a que en cada una de sus etapas se aplican las elecciones más sencillas posibles. El algoritmo inicia con una población generada aleatoriamente. La función de adaptación, es una función matemática para la que se busca el valor óptimo en un determinado intervalo. El algoritmo entra a un ciclo donde el primer paso es una selección de individuos. Esta selección se realiza de tal manera que solo permanezcan los individuos mejor adaptados. Los individuos a cruzar se eligen de forma consecutiva, ya que se supone que el proceso de selección ha reubicado a los individuos de forma eficiente. Se aplica una mutación aleatoria y se determina el nivel de adaptación de la nueva generación de individuos. El criterio de paro, es un número máximo de generaciones en las que no hubo mejora de aptitudes.

El esquema general de un algoritmo genético básico es el siguiente:

---

**Algorithm 2** Algoritmo de un genético básico

---

**Require: Entrada:** Conjunto de datos de entrada  $X$ .

**Ensure: Salida:** Conjunto de los mejores datos aptos para resolver el problema

- 1: *Crear población inicial*
  - 2: *Computar población inicial*
  - 3: **WHILE** *condición de paro no se cumple* **Do**
  - 4: *Selección de individuos para la reproducción*
  - 5: *Cruza de individuos*
  - 6: *Mutación de individuos*
  - 7: *Computar la nueva generación* **END**
- 

La estructura se describe con más detalle a continuación:

#### **Población inicial**

Los individuos de la población inicial suelen ser cadenas de ceros y unos generados de forma completamente aleatoria. Es decir, se va generando cada gen, con una función que devuelve un cero o un uno con igual probabilidad. Es importante dotar al algoritmo genético de población con suficientemente variedad, para poder explorar todas las zonas del espacio de búsqueda.

### Selección de individuos

La idea básica de selección, es utilizar una distribución de probabilidad de selección de una cadena, donde la probabilidad es directamente proporcional a la función de aptitud. Es decir, el proceso de selección debe favorecer la cantidad de copias de los individuos más adaptados. Las técnicas de selección usadas pueden clasificarse en tres grupos: selección proporcional, selección mediante torneo y selección de estado uniforme.

Dos de las técnicas de selección proporcional más utilizadas son la ruleta y sobrante estocástico. Estas se describen a continuación.

#### La Ruleta

Este método ha sido el más utilizado desde los inicios de los AGs. El algoritmo presenta el problema de que el individuo menos apto puede ser seleccionado más de una vez. Sin embargo, su popularidad se debe a su simplicidad. El algoritmo de la Ruleta es el siguiente:

---

**Algorithm 3** Algoritmo de selección (Ruleta)

**Require:** Aptitudes de cada individuo  $f_1, f_2, \dots, f_n$

**Ensure:** Individuos seleccionados

- 1: Calcular la suma de valores esperados  $T$ .
  - 2: Repetir  $N$  veces ( $N$  es el tamaño de la población)
  - 3: Generar un número aleatorio  $r$  entre  $0.0$  y  $T$
  - 4: Ciclar a través de los individuos de la población sumando los valores esperados hasta que la suma sea mayor o igual a  $r$ .
  - 5: El individuo que haga esta suma exceda el límite es el seleccionado.
- 

#### Sobrante Estocástico

El sobrante estocástico reduce los problemas de la ruleta, pero puede causar convergencia prematura al introducir una mayor precisión de selección. La idea principal es asignar determinísticamente las partes enteras de los valores esperados para cada individuo y luego usar otro esquema para la parte fraccionaria. El algoritmo es el siguiente:

---

**Algorithm 4** Algoritmo de selección (Sobranter estocástico)

---

**Require:** Aptitudes de cada individuo  $f_1, f_2, \dots, f_n$ **Ensure:** Individuos seleccionados

- 1: Asignar de manera determinística el conteo de valores esperados a cada individuo (valores enteros)
  - 2: Los valores restantes (sobrantes del redondeo) se usan probabilísticamente para rellenar la población.
- 

**Cruza**

Este es un método de fusión sobre la información genética de dos individuos. Este proceso provee un mecanismo para heredar características a su descendencia donde intervienen ambos padres.

La forma más simple del operador de cruce es el cruce mono punto, que consiste en seleccionar una única posición en la cadena de ambos padres e intercambiar las partes divididas por dicha posición.

**Mutación**

La mutación es un proceso donde el material genético puede ser alterado en forma aleatoria, debidamente a un error en la reproducción o la deformación de genes. A diferencia de la genética humana, la probabilidad en un algoritmo genético es mayor. De hecho en un algoritmo genético, la mutación es una forma de evitar caer en mínimos locales.

La forma más sencilla de mutación consiste en cambiar el valor de una de las posiciones de la cadena. Si el valor es cero pasa a uno, y si es uno pasa a cero.

**Condición de paro**

Es necesario especificar las condiciones en las que el algoritmo deja de evolucionar y se presenta la mejor solución encontrada. La condición de paro más sencilla, se presenta al detectar que la mayor parte de la población ha convergido a una forma similar, careciendo de la suficiente diversidad para que tenga sentido continuar con la evolución.

## 2.6. Técnicas de Evaluación de resultados

Para obtener un modelo y generar predicciones primero dividimos el conjunto de datos en dos partes, un conjunto de entrenamiento y un conjunto de prueba. Las observaciones en el conjunto de entrenamiento se utilizan para construir el modelo deseado. A menudo una pequeña subparte del conjunto de entrenamiento se mantiene con fines de validación y se conoce como conjunto de validación. A veces, un preprocesamiento se realiza normalizando los datos.

Una vez que un modelo es construido, se utiliza para generar predicciones. Las observaciones del conjunto de prueba se mantienen para verificar la exactitud del modelo ajustado obtenido. Si es necesario, una transformación inversa se aplica a los valores pronosticados para convertirlos en la escala original.

A fin de evaluar la exactitud de predicción de un modelo particular o para evaluar y comparar diferentes modelos, se considera el rendimiento relativo en el conjunto de datos de prueba.

Debido a la importancia fundamental de la predicción de series de tiempo en muchas situaciones prácticas, se deben tomar ciertas medidas para seleccionar un modelo. Por esta razón, varias métricas de rendimiento se proponen en la literatura. Cada una de estas medidas es una función de los valores reales y previstos de las series temporales.

En esta Sección se describen algunas métricas de desempeño importantes que son frecuentemente utilizadas por los investigadores, con sus rasgos sobresalientes.

### 2.6.1. Error de Predicción Medio

La métrica es definida como:

$$MFE = \frac{1}{n} \sum_{i=1}^n e_i \quad (2.37)$$

donde  $e_i = y_i - f_i$  representa el error de predicción,  $y_i$  representa el valor de salida real,  $f_i$  es el valor predicho,  $n$  es el tamaño del conjunto de prueba. Las propiedades de la métrica  $MFE$  son:

- Mide la desviación media de los valores pronosticados de los reales.
- Muestra la dirección del error y por lo tanto también se denomina *sesgo de predicción*.

- Los efectos de los errores positivos y negativos se anulan y no hay manera de conocer su cantidad exacta.
- Un MFE cero no significa que los pronósticos sean perfectos o que no contengan ningún error; Sino que sólo indica que las predicciones están en la meta adecuada.
- MFE no penaliza errores extremos.
- Depende de la escala de medida y también se ve afectada por las transformaciones de datos.
- Para un buen pronóstico, es decir, para tener un sesgo mínimo, es deseable que el MFE sea lo más cercano a cero como sea posible.

### 2.6.2. Error de Porcentaje Absoluto Medio

La medida es dada por

$$EPAM = \frac{1}{n} \sum_{i=1}^n \left| \frac{e_i}{y_i} \right| \times 100 \quad (2.38)$$

Las propiedades de la métrica  $EPAM$  son:

- Esta medida representa el porcentaje de error absoluto promedio.
- Es independiente de la escala de medida, pero afectada por la transformación de datos.
- No muestra la dirección del error.
- MAPE no divide las desviaciones extremas.
- En esta medida, los errores opuestos no se compensan entre sí.

### 2.6.3. Error de Porcentaje Medio

La medida es dada por

$$EPM = \frac{1}{n} \sum_{i=1}^n \left( \frac{e_i}{y_i} \right) \times 100 \quad (2.39)$$

Las propiedades de la métrica  $EPM$  son:

- MPE representa el porcentaje de error promedio ocurrido durante la predicción.
- Tiene propiedades similares a EPAM
- Muestra la dirección del error.
- Los errores opuestos se afectan entre sí y se cancelan.
- De la misma forma que MFE, si se obtiene un valor de MPE cercano a cero, no podemos concluir que el modelo haya tenido un rendimiento muy bueno.
- Es deseable que para una buena predicción el MPE obtenido sea pequeño.

#### 2.6.4. Error Cuadrático Medio

La definición matemática está dada por

$$ECM = \frac{1}{n} \sum_{i=1}^n e_i^2 \quad (2.40)$$

Las propiedades de la métrica son:

- Es una medida de la media de la desviación cuadrada de los valores predichos.
- Los errores opuestos no se compensan entre sí, MSE da una idea general del error producido durante la predicción.
- Penaliza los errores extremos ocurridos durante la predicción.
- Es importante señalar que el error total es muy afectado por grandes errores individuales, es decir, los errores grandes son mucho más caros que los errores pequeños.
- No proporciona ninguna idea acerca de la dirección del error general.
- Es sensible al cambio de escala y transformaciones de datos.
- Aunque es una buena medida del error general, no es tan intuitivo y fácilmente interpretable como las otras medidas discutidas anteriormente.

### 2.6.5. Sumatoria del Error Cuadrático

La definición matemática esta dada por

$$ECM = \sum_{i=1}^n e_i^2 \quad (2.41)$$

Las propiedades de la métrica son:

- Mide la desviación cuadrada total de las observaciones, a partir de los valores reales.
- Las propiedades de SEC son las mismas que las de MSE.

### 2.6.6. Error Cuadrático Medio Absoluto

La definición matemática esta dada por

$$ECMA = \frac{1}{n} \sum_{i=1}^n \left( \frac{e_i}{|e_i|} \right) e_i^2 \quad (2.42)$$

Las propiedades de la métrica son:

- Es igual que ECM, excepto que aquí el signo original se mantiene para cada error cuadrático individual.
- ECMA penaliza los errores extremos ocurridos durante la predicción.
- A diferencia de ECM, ECMA también muestra la dirección del error general.
- Los errores positivos y negativos se compensan entre sí.
- Es sensible al cambio de escala y transformaciones de datos.

### 2.6.7. Raiz del Error Cuadrático Medio

La definición matemática esta dada por

$$RECM = \sqrt{ECM} = \sqrt{\frac{1}{n} \sum_{i=1}^n e_i^2} \quad (2.43)$$

Las propiedades de la métrica son:

- Comparte las mismas propiedades del Error Cuadrático Medio

### 2.6.8. Error Cuadrático Medio Normalizado

La definición matemática esta dada por

$$ECMN = \frac{ECM}{\sigma^2} = \frac{1}{\sigma^2} \sum_{i=1}^n e_i^2 \quad (2.44)$$

donde  $\sigma = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2$  y  $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$  Las propiedades de la métrica son:

- Normaliza el MSE obtenido después de dividirlo por la varianza de prueba.
- Es una medida de error balanceada y es muy efectiva para evaluar la precisión de la predicción.
- Cuanto menor sea el valor ECMN, mejor es la predicción.
- Otras propiedades de ECMN son las mismas que las de ECM.



# Capítulo 3

## Metodología

A continuación se presenta una breve reseña de la metodología realizada en este proyecto. El diagrama de la Figura 3.1, representa cada una de las etapas de la metodología. Este diagrama muestra un esquema general de las etapas que se llevaron a cabo durante el proyecto. Cada etapa se analiza de forma independiente, permitiendo comprender a detalle su implementación.

En este Capítulo se describe la metodología y estrategias con que se llevó a cabo la predicción de las tendencias poblacionales. En la tesis se implementó un algoritmo genético para eliminar características que introducen ruido y/o que afectan la precisión y obtener solo aquellas que influyen en una buena predicción. El trabajo comenzó con la búsqueda de bibliografía relacionada con la predicción de series de tiempo.

### 3.1. Conjunto de datos

Debido a la naturaleza del estudio se requirió de una cantidad de datos necesaria para tener una perspectiva clara, por lo menos en lo que concierne a nivel académico que aspectos influyen en él. Los datos debían ser personalizados, de cada municipio con una temporalidad suficientemente cercana que se lograra tener una descripción del comportamiento de sus variables veraz. Es importante destacar que se buscó información de la gran mayoría de los municipios del Estado de México, solo se tomaron los municipios conurbados del estado para este estudio, aunque todos los municipios son diferentes, se vio un comportamiento más peculiar en los que pertenecen a la Zona Metropolitana del Valle de México, pues sus características demográficas y socioculturales son importantemente distintas a las del resto del estado. Los datos se obtuvieron de los bancos de datos del Instituto Nacional de Estadística, Geografía e Informática (INEGI), algunos de los que se registran pertenecen a los censos y conteos poblacionales, así como a encuestas aplicadas a comercios o mediciones de variables macroeconómicas

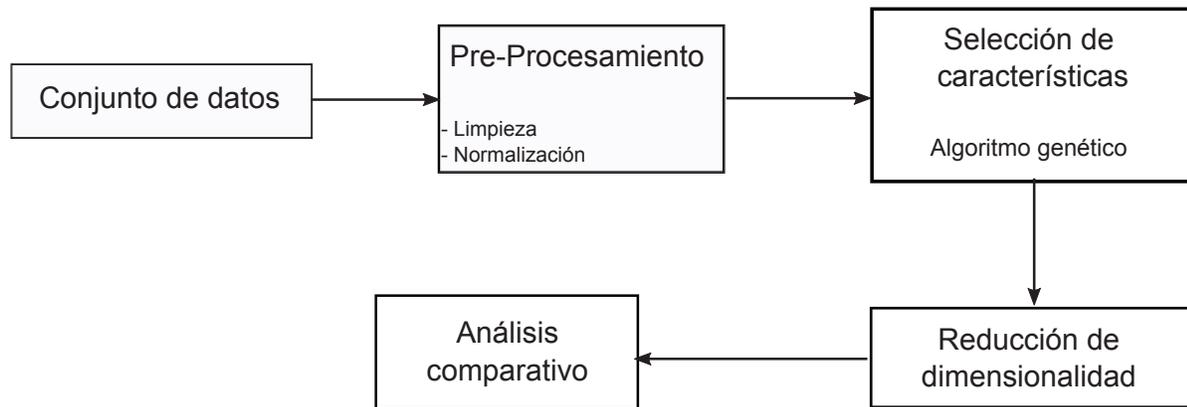


FIGURA 3.1: Metodología

principalmente del Banco de Información Económica (BIE) de INEGI, dependiendo de su naturaleza la información procede de diferentes fuentes. Fue recaudada la mayoría de Información en un lapso de varios meses, la búsqueda de información arroja resultados que son útiles y otros que no lo son, de forma que el proceso discriminatorio se vuelve meticuloso y en muchas ocasiones repetitivo. Ya que se da en diferentes etapas en este proceso, según se necesita la información, o según se ve logren los resultados deseados, de otra forma se debe buscar más información; Este proceso provoca que la búsqueda sea continua. Cuando las variables no satisfacen los objetivos deseados se cicla el proceso hasta que se logra tener la información deseada. El proceso puede describirse de manera sencilla de la siguiente forma: recolección de datos, evaluación de variables, búsqueda de variables faltantes de acuerdo al paso anterior, procesamiento de datos, aplicación en modelos prueba, delimitación de fechas de estudio, búsqueda de variables faltantes o de apoyo, aplicación a modelos y modelado final.

### 3.1.1. Variables utilizadas

Las variables utilizadas son de diferentes criterios, hay variables económicas, sociales, demográficas, actividades agropecuarias e industriales, entre otras muchas.

Las variables un primero plano se consideró fueran abundantes para abarcar la mayor cantidad de información posible para ver qué carteles son significativas para investigación, los procesos se filtrado y análisis de datos van reduciendo la cantidad de las variables. Uno de los filtros a los que se enfrentan las variables es a la relación "directa" que tienen con el objeto de estudio que es el académico, no obstante, aunque tienen prioridad las que se relacionas más directamente, como puede ser PIB destinado a la

educación, las demás variables se introducirán el ejercicio como una forma de medir si intervención en el objetivo, en ocasiones sorprende ver variables poco relacionadas con el objetivo principal que estadísticamente tienen una correlación positiva entre sí.

Después el filtro de la temporalidad, la información debe ser homogénea, significa que debe de coincidir en rubros como el tiempo en que se obtuvo dicha medición, la unidad de medida con que se presenta la información, que se refieran a la misma región geográficas, entre otras. La temporalidad se refiere al primer caso, que coincida el tiempo que se usó para medir determinado dato. Podríamos decir que es fundamental saber en qué tiempo sucedió cierta medición en todos municipios estudiados, de otra forma que estemos evaluando el mismo dato de todos los municipios en el mismo tiempo y no en diferentes tiempos aunque hablemos de la misma variable; no sería útil medir la Población Económicamente Activa (PEA) en Ecatepec en el año 2007 y Compostela con la de Coacalco en el año 1990, pues si quisiéramos comparar el porcentaje de la PEA de ambos municipios nos resultaría bastante disipar el resultado ya que estamos hablando de tiempos muy distintos. Es por eso que la homogeneidad en la temporalidad de los datos es fundamental. También tenemos la disponibilidad de datos dentro del intervalo de tiempo en que se desarrolla el estudio. No toda la información está disponible para todos los municipios, hay municipios que cuentan con información que otros no, por lo tanto si deseáramos hacer la comparación de ambos municipios confrontando sus variables tendríamos un problema al darnos cuenta de que uno de ellos tiene una variable que el resto no tiene, ¿con que información se va a comparar si no existe?, por lo tanto, dentro del procesamiento y tratamiento de datos se eliminarán aquellas variables que sean exclusivas de unos cuantos municipios y no de todos en general, así garantizaremos que el modelo que se aplique tendrá la información en cada municipio.

Cuando se ve que cierta información no es útil, se desecha, pero, cuando hay información que haga falta entonces se deberá buscar y agregar al conjunto de datos, este proceso se puede dar varias veces hasta que se ha determinado un buen conjunto de datos que sea consistente y homogéneo.

Después de estos procesos entrará el filtrado de datos por medio de aplicación de modelos de Inteligencia Artificial que determinarán la significancia de las variables, las variables generales que se introdujeron en la investigación sin haber aplicado este último filtrado de datos fueron 931 y son las siguientes:

### 3.1.2. Municipios Obtenidos

Los municipios estudiados son los de la Zona Metropolitana del Valle de México, son 20: Acolman Atizapán de Zaragoza Coacalco de Berriozábal Cuautitlán Chalco Chicoloapan Chimalhuacán Ecatepec de Morelos Huixquilucan Ixtapaluca Naucalpan de Juárez Nezahualcóyotl Nicolás Romero La Paz Tecámac Tepetzotlán Texcoco Tlalnepantla de Baz Tultitlán Cuautitlán Izcalli

## 3.2. Pre-Procesamiento

El preprocesamiento de datos es un paso muy importante durante el proceso llevado a cabo. Se trata de diversos tipos de procesamiento que se realizan con los datos brutos para transformarlos en datos que tengan formatos que sean más fáciles de utilizar.

La mayoría de la veces, los datos frecuentemente no están limpios, faltan valores clave, algunos datos contienen inconsistencias y en ocasiones suelen contener ruido, ya sea por variables con errores o valores atípicos. Sin un preprocesamiento de datos, estos errores en los datos sobrevivirían y disminuirían la precisión. La principal importancia de este paso radica en que mejora la calidad de los datos permitiendo aumentar el desempeño de clasificación y/o predicción.

### 3.2.1. Limpieza

La limpieza de datos consiste en detectar los datos erróneos o irrelevantes y descartarlos. Una de las actividades dentro de la limpieza de datos es el tratamiento de datos ausentes. Esto sucede cuanto falta el valor de un atributo. Para rellenar este valor se pueden tomar diversas estrategias, algunas de las cuales son: utilizar la media o la moda de los valores anteriores y posteriores al dato, generar un valor aleatorio basándose en una distribución gaussiana, y en otros casos utilizar algún método de interpolación.

La falta de limpieza adecuada en los datos es el principal problema en clasificación y/o predicción. Algunos de las tareas de preprocesamiento de datos son las siguientes:

1. Rellenar valores faltantes
2. Identificar y eliminar datos que se pueden considerar un ruido
3. Resolver redundancia

#### 4. Corregir inconsistencias

En los experimentos realizados La limpieza de los datos se realizo utilizando dos técnicas. Primero se eliminaron aquellas variables en las que en ningun año presentaban valores y cuando alguna variable presentaba valores, pero en algun año habia alguna ausencia se rellenaron manualmente los valores perdidos utilizando interpolación.

### 3.2.2. Normalización de datos

La normalización es empleada cuando se tienen atributos con órdenes de magnitud muy diferentes, en nuestro caso se presenta en muchas características de nuestro conjunto de datos. La normalización evita que los atributos con valores muy altos ganen un peso significativo en el modelo final que aquellos con valores más bajos.

Para evitar esto, todas las características extraídas fueron normalizadas con la siguiente relación:

$$f_{ij} = \frac{T_{ij} - \mu_j}{\sigma_j} \quad (3.1)$$

donde  $i = 1, \dots, m$  y  $j = 1, \dots, n$ ,  $\mu_j$  y  $\sigma_j$  representan la media y desviación estándar de la  $j$  -ésima característica,  $T_{ij}$  representa la  $j$  -ésima característica del  $i$  -ésimo vector,  $m$  es el número de datos y  $n$  el número de características. Las características normalizadas tienen media cero y desviación estándar igual a 1.

### 3.3. Cross-validation

La validación cruzada (*Cross-validation*), es un método para evaluar y comparar los algoritmos de aprendizaje mediante la división de datos en dos segmentos. El primero se utiliza para entrenar el modelo y el segundo para validar el modelo.

El algoritmo primero divide los datos en  $k$  partes iguales. Después realiza  $k$  iteraciones de entrenamiento, tomando en cada iteración como conjunto de prueba un subconjunto diferente y construyendo el modelo con los subconjuntos restantes. El índice de error estimado es la media de todos los errores obtenidos en cada entrenamiento.

La ventaja de evaluar a partir de  $k$  combinaciones de datos de entrenamiento y prueba hace que el método sea más preciso. Sin embargo, en una evaluación con un valor

alto en  $k$  el proceso es lento al momento de computar. La elección del número de iteraciones depende de la medida del conjunto de datos, aunque lo más común es utilizar la validación cruzada de 10 iteraciones.

Se utilizó validación cruzada durante los procesos de selección de características y evaluación del modelo.

### 3.4. Selección de características

Un problema más difícil es la eliminación de los datos ruidosos. Estos casos corresponden con ejemplos que son significativamente diferentes o son inconsistentes con el conjunto de datos. El tratamiento de los datos ruidosos es una de las tareas más difíciles. Para detectar qué es un dato ruidoso y qué no, pueden utilizarse diversas estrategias, basándose algunas de ellas en umbrales de probabilidad de determinadas distribuciones de probabilidad. Muchos algoritmos de minería de datos tratan de minimizar el impacto de los datos ruidosos en el modelo final, o incluso eliminarlo. Los datos ruidosos pueden provenir de errores en los aparatos de medidas, de medidas reales y correctas pero altamente improbables, o de casos excepcionales. En cualquier caso, las características que introduzcan ruido en el modelo deben ser eliminadas con mucha precaución.

En esta fase del preprocesado de datos, se descartan atributos que no son relevantes para la toma de decisión. En general, el volumen de datos original suele exceder de lo deseable y de lo práctico para su aplicación. Además está bien estudiado que gran parte de la información es redundante, principalmente debido a que muchas variables están correlacionadas. Por lo tanto, una importante reducción de atributos puede aplicarse a los datos sin pérdida significativa de información.

El objetivo de la selección de atributo es encontrar el conjunto mínimo de atributos de forma que la distribución resultante de probabilidad de las clases de datos es tan próxima como sea posible a la distribución original usando todos los atributos.

Para  $N$  atributos hay  $2^N$  posibles subconjuntos. Por lo tanto una búsqueda exhaustiva del subconjunto óptimo es computacionalmente costoso si  $N$  es grande. Por lo tanto, son necesarios métodos heurísticos que exploren el espacio de soluciones. Aunque existen en la literatura diversos algoritmos para la selección de características, en esta tesis se utilizó un algoritmo genético básico para seleccionar las mejores características.

### 3.4.1. Reducción de la Dimensionalidad

En la reducción de la dimensionalidad se aplica una transformación para obtener una representación reducida o comprimida de los datos originales. Si los datos originales pueden ser reconstruidos desde los datos comprimidos sin pérdida de información, entonces la reducción se denomina sin pérdida de datos. Por el contrario, si se puede reconstruir los datos solo de forma aproximada entonces se denomina reducción de dimensionalidad con pérdida de datos. Un ejemplo muy popular de reducción de la dimensionalidad es el análisis de componentes principales.

## 3.5. Predicción

En esta parte se realiza la comparación de los modelos que se han utilizado en la investigación, se realizaron análisis de dos clases de graficas unas que llevan la predicción a los últimos cinco años del intervalo de estudio, para apreciar los comportamientos de cada uno de ellos, y damos compararlos entre sí: La otra parte de los modelos, se analiza con un periodo de predicción mayor con la finalidad de poder comparar los resultados pronosticados con los reales.

Los modelos que se ocuparon para realizar las predicciones de los datos son los siguientes:

1. Regresión Lineal
2. Redes Neuronales Artificiales
3. Máquinas de Vectores Soporte con kernel Gaussiano
4. Máquinas de Vectores Soporte con kernel Polinomial
5. Máquinas de Vectores Soporte con kernel RBF

## 3.6. Análisis comparativo de resultados

En esta Sección se llevará a cabo un análisis comparativo entre los datos reales y los pronosticados por los diferentes modelos, para poder así medir su ajuste a los datos reales, el objetivo es lograr ver cuál es el mejor modelo que mejor se ajusta alzo datos reales, se presentaran datos e información de algunas métricas indicadoras basadas en error para realizar las comparaciones. En esta sección para este estudio, únicamente se

presenta uno de los municipios estudiados, Tlalnepantla, para hacer el análisis completo de su precisión, se tomaron en cuenta los últimos cinco años de donde se tienen datos reales, 2007 a 2011, para estudiar. Se han de aplicar estudios a todos los niveles de escolaridad, Pre-Primaria, Primaria, Secundaria, Bachillerato y Profesional-Técnico, en Tlalnepantla, las métricas también se aplican todas en los niveles mencionados. Se ha determinado que por causa del espacio únicamente se presentara el municipio de Tlalnepantla en el estudio.

## Capítulo 4

# Resultados experimentales

En la presente sección se realiza la comparación de los modelos que se han utilizado en la investigación, se ha de realizar un análisis de dos clases de graficas unas que llevan la predicción a los últimos cinco años del intervalo de estudio, para que miremos nosotros cuales son los comportamientos de cada uno de ellos, y damos compararlos entre sí: La otra arte de los modelos, se analiza con un periodo de predicción mayor con la finalidad de poder comparar los resultados pronosticados con los reales.

Los municipios del Estado de México que pertenecen al área conurbada son los siguientes:

1. Acolman
2. Atizapán
3. Chalco
4. Chicoloapan
5. Chimalhuacán
6. Coacalco
7. Cuautitlán
8. Ecatepec
9. Huixquilucan
10. Ixtapaluca
11. La Paz
12. Naucalpan

13. Nezahualcóyotl
14. Nicolás Romero
15. Tecámac
16. Tepotzotlán
17. Texcoco
18. Tlalnepantla
19. Tultitlan

De los municipios anteriores los siguientes son los que tendrán el análisis más detallado en cuanto a un intervalo de predicción más larga:

1. Ecatepec
2. Naucalpan
3. Nezahualcóyotl
4. Texcoco
5. Tlalnepantla

Análisis por comparación entre modelos

Se va a analizar las gráficas de los municipios, los modelos que se ocuparon para realizar las predicciones de los datos son los siguientes:

1. Regresión Lineal
2. Redes Neuronales Artificiales
3. Máquinas de Vectores Soporte con kernel Gaussiano
4. Máquinas de Vectores Soporte con kernel Polinomial
5. Máquinas de Vectores Soporte con kernel RBF

En los municipios se compararán estos modelos para mirar su comportamiento entre ellos.

## 4.1. Acolman

La grafica 4.1 que podemos mirar muestra el histórico que los datos de personas que han egresado de los diferentes niveles educativos, desde 1994 al 2011, posteriormente aparecen los datos predichos para los siguientes cinco años, en la gráfica presente podemos contemplar que los datos más altos hasta antes del 2005 corresponden a los grados de primaria, a partir del 2005 incrementa notablemente el egreso de estudiantes de preescolar, dato que coincide con el decreto de escuela preescolar obligatoria para todos los niños.

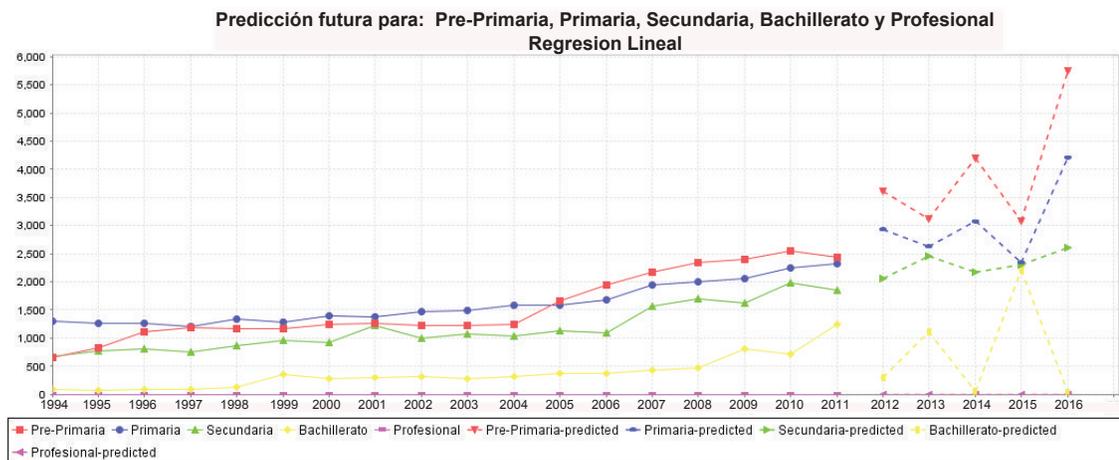


FIGURA 4.1: Resultados obtenidos utilizando Regresión lineal

Podemos nosotros también identificar que entre más alto es el nivel educativo, menos estudiantes egresan; Mientras primaria y preescolar dominan en diferentes periodos de tiempo, y sus graficas son relativamente parecidas o cercanas, de allí cada nivel educativo presenta una cada vez mayor distancia al número de egresados del nivel académico anterior. Tenemos que primaria y preescolar cuentan con el número más alto de egresados la secundaria presenta mientras tanto, niveles inferiores a los de sus grados previos sobre todo a partir del 2005. Con toda la situación empeora reduciéndose a menos del 50 % de egresados de bachillerato en comparación con los egresados de sus grados académicos previos, la brecha concluye en el egreso de profesionistas, que no tiene punto de comparación con ninguno de los gráficos de los demás niveles educativos. Es destacable también entender que los egresados es una variable con características que daremos a conocer, un egresado no es cada estudiante inscrito en el nivel académico respectivo, tampoco es el numero esperado de alumnos que concluyan el nivel, se refiere al número de estudiantes que han terminado por completo cierto grado

académico y han quedado registrados como alumnos que oficialmente terminaron un plan de estudios, egresados. De un 100 % de ingresos, alumnos inscritos, el porcentaje que termina es menor y va desde bajísimos hasta altísimos niveles de deserción escolar, alumnos que no concluyen sus niveles educativos. El histórico es igual en todas las gráficas, la parte que queremos comparar son los años de 2012 al 2016 entre los diferentes modelos. Podemos observar que los pronósticos del modelo de Regresión Lineal (RL) y el modelo SVM con kernel Pol. son similares en varios de los datos predichos, salvo en el año 2013 para Secundaria donde su comportamiento es opuesto en ambos modelos.

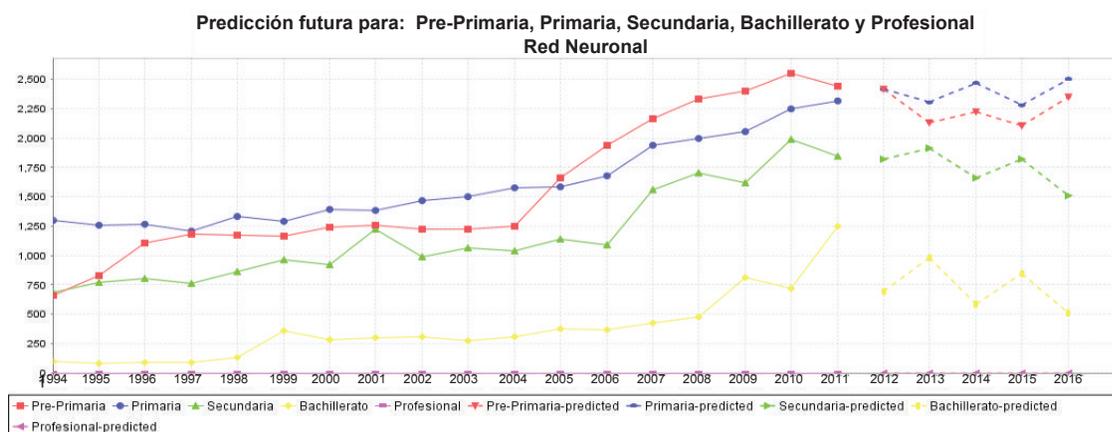


FIGURA 4.2: Resultados obtenidos utilizando redes neuronales

Los otros dos modelos SVM, tanto Gauss como RBF muestran un comportamiento casi lineal y con pendiente negativa en el transcurso de los años, que denota un comportamiento decreciente a diferencia de la línea de tendencia general cuyo comportamiento es creciente. Para terminar, el modelo de Redes Neuronales Artificiales (RNA) Fig. 4.2 muestra un comportamiento más apegado al patrón que viene dándose en los años anteriores, posiciona el pronóstico de Primaria por encima del Pre Primaria como al comienzo del periodo de estudio, cabe señalar que es la única proyección que arroja estos datos. Es notorio que el modelo RNA también es el que arroja datos que denotan una mayor continuidad respecto al último año del histórico, pues otros modelos como el SVM Gauss y el RL generan en algunas variables saltos grandes que se separan considerablemente de la tendencia que llevaba. Incluso también notamos que entre más pasa el tiempo en estos modelos los datos se empiezan a comportar de manera radical con saltos bastante pronunciados que en datos reales no se presentan. Aunque el análisis es del comportamiento de cómo se modelan los datos futuros y la descripción de ellos es en ocasiones difícil de creer, puede ser en la realidad que sean datos confiables.

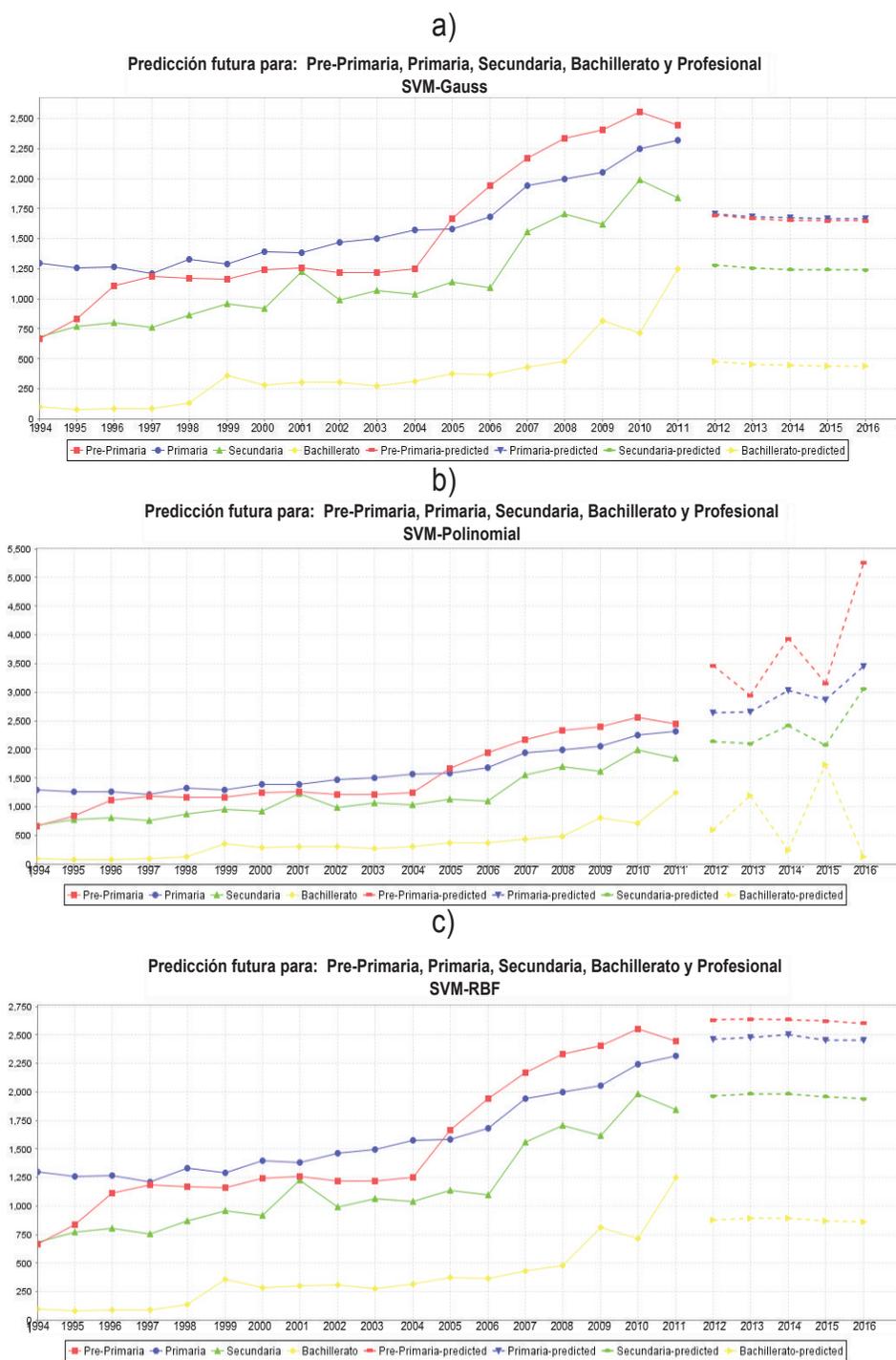


FIGURA 4.3: Resultados obtenidos utilizando SVM con diferentes Kernels  
a) Gauss b) Polinomial c) RBF

## 4.2. Atizapan

Miramos en las Figura de Atizapán 4.4 patrones diferentes a los de Acolman, Atizapán muestra un comportamiento más uniforme en primaria, e incluso a largo plazo decreciente; Pre Primaria muestra un importante incremento para 2005, con un aumento considerable en el número de egresados del nivel académico y que además se mantiene en adelante a partir de dicho año, Secundaria por su lado tiene un crecimiento irregular pero constante. Profesional-Técnico muestra un punto máximo en el año 2000 de allí en adelante lleva una tendencia decreciente lenta, Bachillerato por su lado es creciente en el periodo de estudio. Nuevamente los modelos de RL y SVM Polinomial muestran comportamientos muy similares en los periodos de predicción, sobre todo en la proyección de Pre Primaria que tiene un valor mayor en 2012 y cae en 2014, este último comportamiento también lo presenta el modelo de RNA (Figura 4.4 b)) en 2014 solo que en 2012 el valor de la misma variable es menos pronunciado que el de los otros dos modelos. Así mismo los tres modelos antes mencionados, tienen el mismo patrón de comportamiento en Profesional-Técnico con un valor mayor en 2014, máximo local entre 2012 y 2016; SVM Gauss y SVM RBF (Figura 4.4 c) y e)) cambian en esta variable, mientras Gauss presenta su máximo en 2015, RBF muestra un comportamiento casi lineal, pero de pendiente negativa.

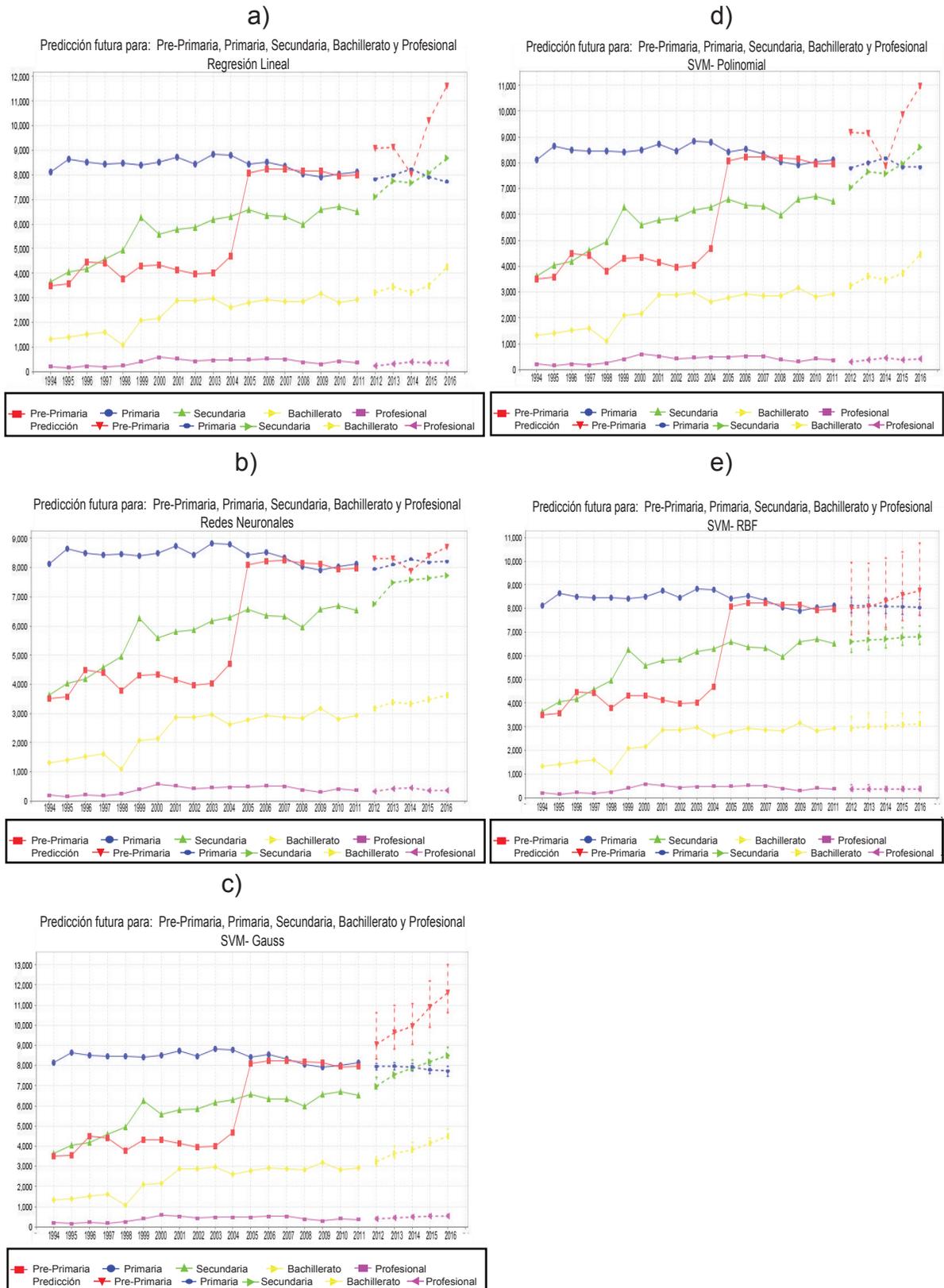


FIGURA 4.4: Resultados obtenidos utilizando diferentes métodos de regresion a) Regresión lineal b) Redes neuronales c) SVM método de Gauss d) SVM-Polinomial e) SVM-RBF

En cuanto al resto de variables, tanto SVM Gauss como SVM RBF tienen tendencias y patrones similares ciertamente las de Gauss son más pronunciadas que RBF, los comportamientos de ambos modelos tienen una tendencia lineal con igual signo en la pendiente, solo menor inclinación en RFB. El igual que en Acolman, las líneas que describen el grado Profesional-Técnico varían muchísimo del resto de variables, pudiendo decir que, si el nivel de Bachillerato tiene menos egresados por debajo del 50 % de los niveles de Pre Primaria y Primaria, Profesional-Técnico se encuentra por debajo del 10

### 4.3. Chalco

Chalco, tiene un comportamiento interesante (Figura 4.5), ya que sus cifras en 1994 desde Pre Primaria hasta y Primaria son más altas que el resto de datos reales, y caen durante los dos años siguientes tocando su punto mínimo para volver a empezar un despegue lento en todo el periodo de estudio, aun con todo no alcanzan su nivel inicial, comportamiento similar en Secundaria donde hay puntos altos en el 94 que disminuyen hacia el 95 y 96 tocando su punto más bajo, solo que en esta ocasión Secundaria en 2005 logra superar su punto inicial del 94. Vemos que en el 96 para los tres primeros niveles educativos hay un importante impacto negativo en los números de egresados, que de manera inmediata comienza su crecimiento otra vez para el siguiente año. La Secundaria además tiene un comportamiento con altas y bajas hacia los años subsecuentes, mostrando un patrón constante pero menguante hacia los años futuros, de oscilamiento en sus números. Por su parte, Bachillerato tiene un comportamiento creciente, podemos decir que es el de mejor o más favorable comportamiento, al igual que Secundaria tiene un patrón de oscilamiento, altos y bajos durante el trayecto de todos los años reales. Profesional-Técnico tiene un comportamiento menos favorable, con un año favorable en 2001 que salta de la línea de tendencia para volver a caer el siguiente año, su comportamiento es generalmente lineal, aun en el último periodo real - 2011 - con crecimiento negativo. La brecha entre Bachillerato y Profesional-Técnico es con los años cada vez mayor, lo que se puede interpretar como un índice de gente que egresa del bachillerato, pero ya no ingresa a la universidad, o por lo menos no la termina.

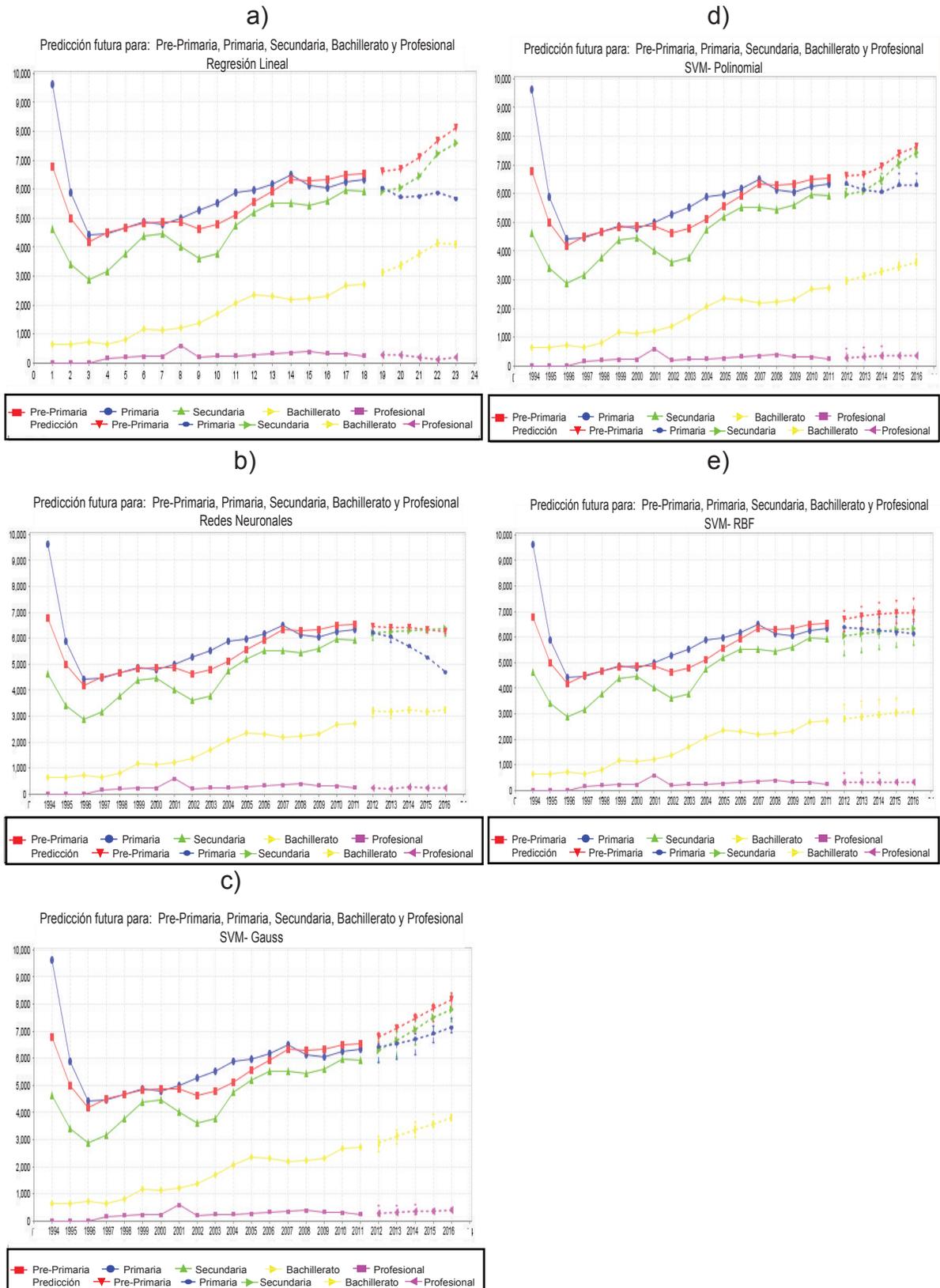


FIGURA 4.5: Resultados obtenidos utilizando diferentes métodos de regresión a) Regresión lineal b) Redes neuronales c) SVM método de Gauss d) SVM-Polinomial e) SVM-RBF

Por último, en las predicciones, nuevamente las figuras entre SVM Polinomial (Figura 4.5 d)) y RL (Figura 4.5 a)) se ven parecidas salvo por un cambio de rotación y desfasamiento vertical en Primaria las tendencias de ambos modelos son similares, Pre Primaria y Secundaria con una tendencia positiva, Profesional-Técnico y Primaria con una negativa, solo Bachillerato tiene diferencias más definidas en los últimos dos periodos de predicción, sobre todo por el año penúltimo que en RL es mayor que el último, y sucede en sentido contrario en SVM Pol, proveyendo tendencia negativa en RL al final, y positiva en SVM Pol. SVM Gauss y SVM RBF (Figura 4.5 c) y e)), sus comportamientos predichos son cercanos a la linealidad. Gauss predice tendencias más crecientes a la larga que RBF, donde vemos que las tendencias son menos crecientes, incluso negativos para Primaria. Por otro lado, Profesional-Técnico tiene un comportamiento más uniforme en el periodo de predicho respecto a las demás variables que muestran comportamientos más variables, esto se debe muy probablemente a la capacidad máxima de recepción de estudiantes en Universidades.

El modelo RNA en contraste con los demás muestra una caída creciente de la variable Primaria a lo largo de los años de predicción, Pre Primaria también resulta con pendiente negativa, el resto parece ser uniforme en el resto de los años.

#### 4.4. Chicoloapan

Miramos en la Figura 4.6 que los índices son crecientes dentro del periodo real, lo que podemos observar es un crecimiento uniforme, sin grandes saltos, ni caídas importantes. Profesional-Técnico sigue siendo una línea muy por debajo de las demás y con tendencias también negativas al largo plazo. Encontramos predicciones interesantes, en las gráficas de varias de ellas hay picos, según el modelo de RNA Bachillerato tiene tendencia creciente con muy buenos resultados al lapso de los 5 años predichos, Primaria, Pre Primaria y Secundaria tienen comportamientos crecientes con algunos puntos de crecimiento de salto al final la terminan con pérdidas, es interesante porque los modelos de RNA se basan en el aprendizaje de patrones de comportamiento y a simple vista no se ven saltos tan pronunciados, ni caídas tan bruscas como en las gráficas de los datos predichos.

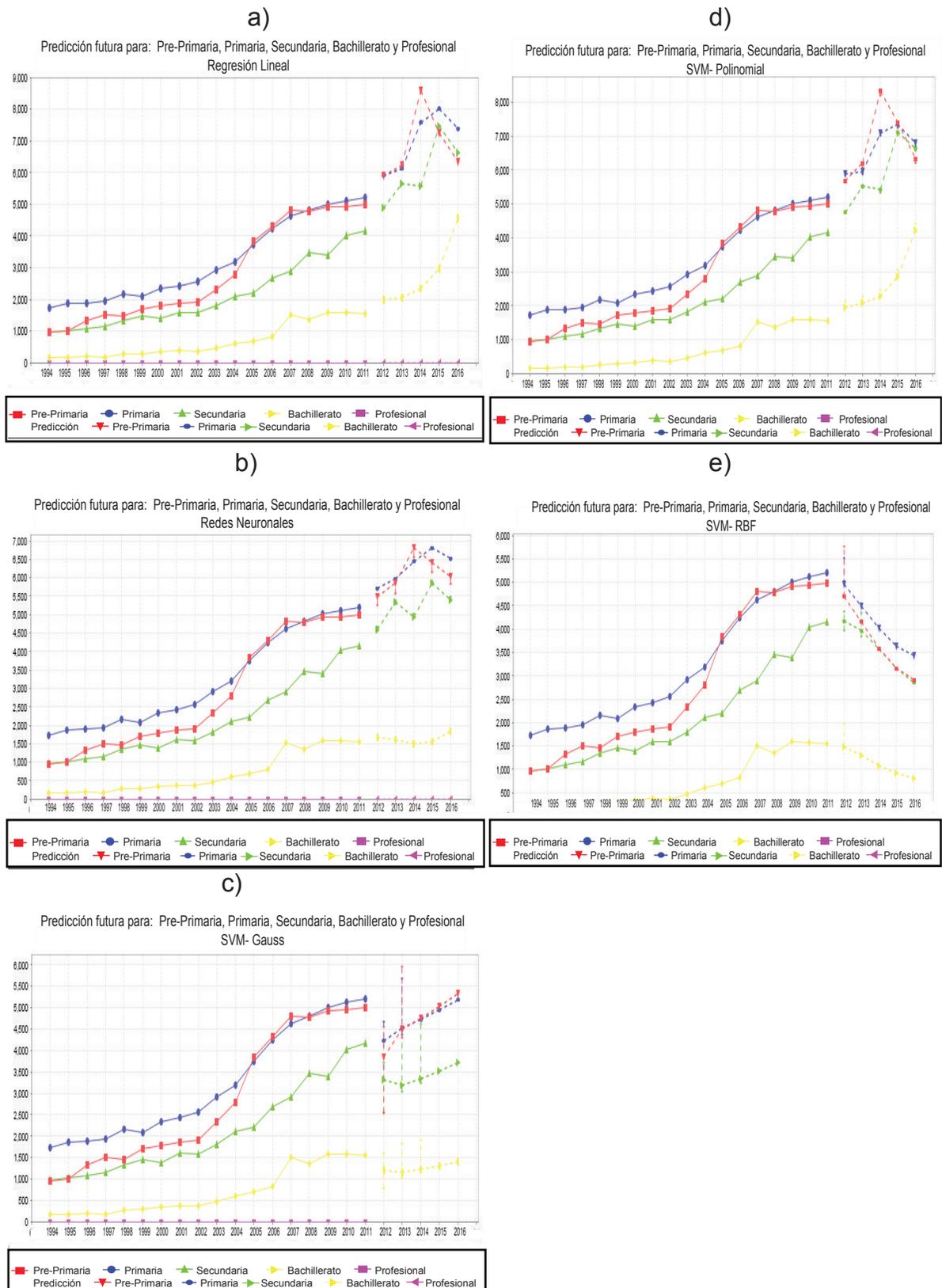


FIGURA 4.6: Resultados obtenidos utilizando diferentes métodos de regresion a) Regresión lineal b) Redes neuronales c) SVM método de Gauss d) SVM-Polinomial e) SVM-RBF

Respecto a las Predicciones, como miramos en la Figura 4.6 d) y a) vemos similitudes en las líneas de predicciones entre SVM Pol. y RL, ambas se comportan con picos en Pre Primaria seguidas de un descenso de dos años, el modelo RNA tiene un pico también en Pre Primaria, aunque no tan alto como en los modelos anteriores, muestra también el un máximo en Primaria en el año 2015, Secundaria es muy similar en los tres modelos en cuanto al comportamiento, el máximo en 2015 en los tres modelos es menor en RNA, aun con todo es muy similar en comportamiento. Bachillerato es parecida en SVM Pol. y RL con un crecimiento exponencial, no es el caso del modelo RNA, donde el comportamiento también es una curva, peor más acostada, dando lugar a un valle. Profesional-Técnico es básicamente imperceptible comparándolo con el resto de los modelos. Posteriormente tenemos los modelos SVM con Kerneles Gauss y RBF (Figura Figura 4.6 c) y e)), en contraste con los municipios estudiados previamente aquí varían mucho más entre ellos, principalmente en la polaridad de la pendiente, mientras en el modelo Gauss las pendientes de las líneas de predicción de las variables son positivas en su mayoría, en RBF no es así, son negativas. Un punto observable es que el paso de los datos reales a los predichos es decir del 2011 al 2012 RBF tiene un cambio menor para comenzar a caer en los años siguientes, mientras que en Gauss hay una caída importante para comenzar enseguida a subir.

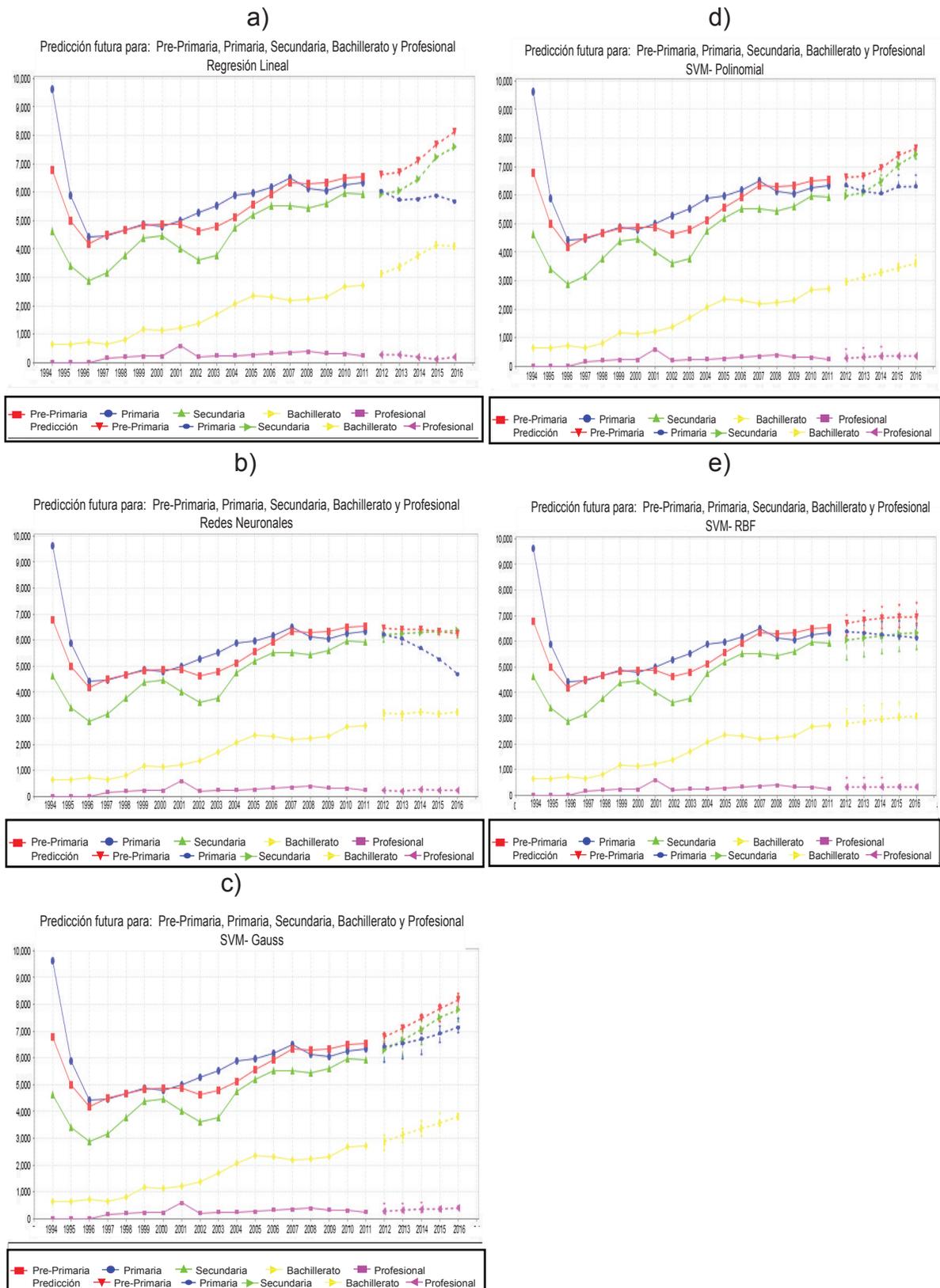


FIGURA 4.7: Resultados obtenidos utilizando diferentes métodos de regresion a) Regresión lineal b) Redes neuronales c) SVM método de Gauss d) SVM-Polinomial e) SVM-RBF

## 4.5. Chimalhuacan

En el municipio de Chimalhuacán, podemos observar, según la Figura 4.7 comportamientos crecientes en general, Pre Primaria tiene un periodo de estancamiento entre 1997 y 2003 para posteriormente comenzar un ascenso entre 2004 y 2005, enseguida sigue creciendo con una tasa relativamente uniforme durante el resto de los periodos de estudio. Primaria por su parte muestra un periodo de recesión en 2006 del que no se recupera dentro del periodo de los datos reales, precedido por un continuo lapso de crecimiento desde el inicio del periodo de estudio. Secundaria también muestra crecimiento en su línea de tendencia, con un periodo de estancamiento de 2007 a 2011 Bachillerato no deja de crecer, aunque a una tasa inferior al resto de las variables ya mencionadas. Profesional-Técnico coincide con Bachillerato en el primer año del periodo de estudio, podemos asumir que los estudiantes que egresaban del bachillerato eran los que continuaban y terminaban sus estudios profesionales, la brecha entre estas dos variables se empieza a abrir, para 1995 Profesional-Técnico muestra un retroceso para mantenerse de allí en adelante sin cambio, mientras lentamente Bachillerato comienza a despegar.

En cuanto a las predicciones, tenemos el ejemplo de la Figura 4.7 que en Pre Primaria: RL, SVM Gauss, SVM Pol. (Figura 4.7 a) c) y d)), muestran crecimiento con un rápido despegue, RNA también muestra crecimiento pronunciado en esta variable, el resto de variables no muestran el mismo comportamiento que Pre Primaria, solo RBF muestra un comportamiento negativo para la variable en los años de estudio predicho. Tenemos en Primaria que en tres modelos RNA, Gauss y RL hay un comportamiento de “cuna” ligera, los otros dos modelos muestran también un comportamiento similar, pero mucho más ligero, casi se ven lineales al final se ven muy similares las predicciones, después del periodo de predicción parece que todas las gráficas muestran caída, a excepción de SVM Pol y RNA, sin embargo, SVM Pol tiene un último periodo negativo, lo que nos deja solo a RNA con una tendencia positiva en Primaria. Secundaria también tiene un comportamiento parecido al de la Primaria, sus graficas muestran en cuatro de los modelos excluyendo a SVM RBF, una cuneta o valle esta ocasión más pronunciada que el de la primaria, solo que, con diferentes pendientes, LR, SVM Gauss son casi iguales, SVM Pol se distingue por que, aun guardando la forma similar al resto de las gráficas, tiene una pendiente más alta, teniendo entre cada uno de sus años crecimiento. Bachillerato nuevamente lleva forma de cuneta en RNA, RL y Gauss, en SVM Pol se aprecia una línea variante y en SVM RBF una línea casi recta. Por su parte Profesional-Técnico no tiene gran diferencia entre los cinco modelos, todos lo muestran

con un comportamiento casi lineal.

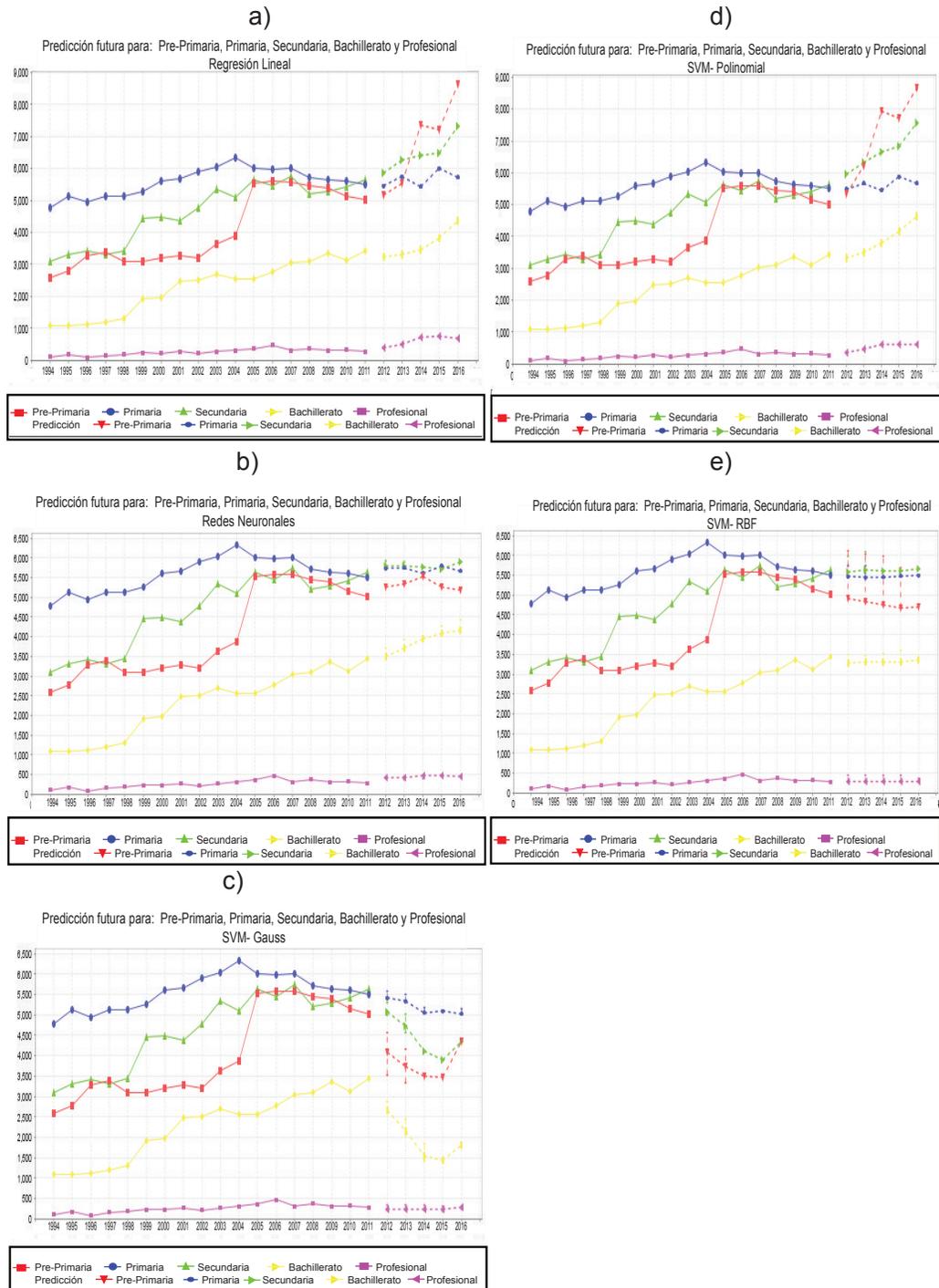


FIGURA 4.8: Resultados obtenidos utilizando diferentes métodos de regresion a) Regresión linealr b) Redes neuronales c) SVM método de Gauss d) SVM-Polinomial e) SVM-RBF

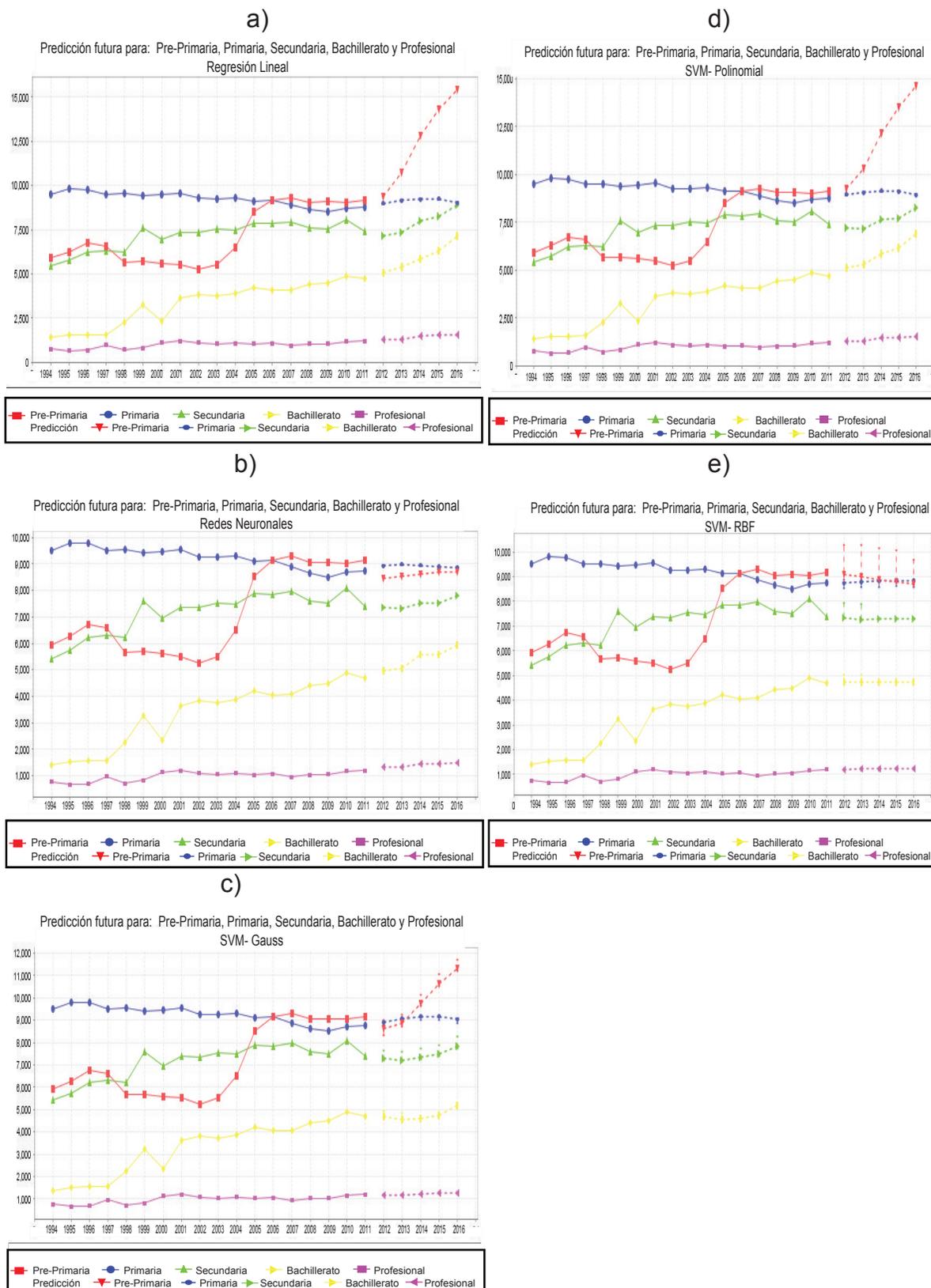


FIGURA 4.9: Resultados obtenidos utilizando diferentes métodos de regresión a) Regresión lineal b) Redes neuronales c) SVM método de Gauss d) SVM-Polinomial e) SVM-RBF

## 4.6. Coacalco

Según la Figura 4.8 podemos ver que en Pre Primaria y en Primaria los datos se pueden apreciar con puntos máximos en 2006 y 2004 respectivamente para comenzar a descender continuamente, los comportamientos denotan un punto en el que se pudo lograr buen incremento en egresados, sobre todo en Pre Primaria del 2004 al 2005 donde tiene un salto importante, sin embargo, es un número que no se pudo sostener, lentamente, pero continua es su caída.

Por otra parte, secundaria muestra en el mismo periodo de descenso de las dos variables previas un estancamiento, presenta su máximo en 2007 sin embargo, de allí el siguiente máximo más cercano a 2007 es 2011 que aun con todo no alcanza 2007, posicionándose en 2011 aun sobre primaria y Pre Primaria, situación destacable en este municipio, ya que generalmente en los municipios estudiados se ha observado que Primaria y Pre Primaria son los niveles académicos con mayor número de egresados. Bachillerato tiene un comportamiento continuamente creciente, años con pequeñas caídas, pero creciente, Profesional-Técnico presenta también un máximo en 2006 que posteriormente al igual que el resto de las variables cae en el siguiente intervalo de años hasta el año 2011. Se destaca que los puntos máximos para la mayoría de variables se dan entre los años de 2004 y 2007.

Como nos ha acontecido en municipios anteriores podemos ver que los modelos SVM Pol y RL tienen comportamientos muy parecidos tendencias similares, picos y valles similares, SVM RBF por otra parte no muestra grandes variaciones en sus predicciones, parecen todas líneas rectas con diferentes pendientes. Gauss nos dice que habrá una recesión seguido de un crecimiento en los años de predicción de la mayoría de las variables. Por ultimo RNA describe secundaria como la variable con mejores cifras a excepción de 2015 donde tiene una pequeña diferencia con Primaria que la supera por pocos puntos; También considera que Primaria estará sobre Pre Primaria acercándose a ella en 2014 únicamente para volverse a separar, Bachillerato es el único modelo nos dice tendrá un crecimiento logarítmico no tan pronunciado, pero con dicha forma. Por ultimo nos presenta un escenario más optimista para Profesional-Técnico que el año 2011, último año con datos reales, pero con un comportamiento uniforme.

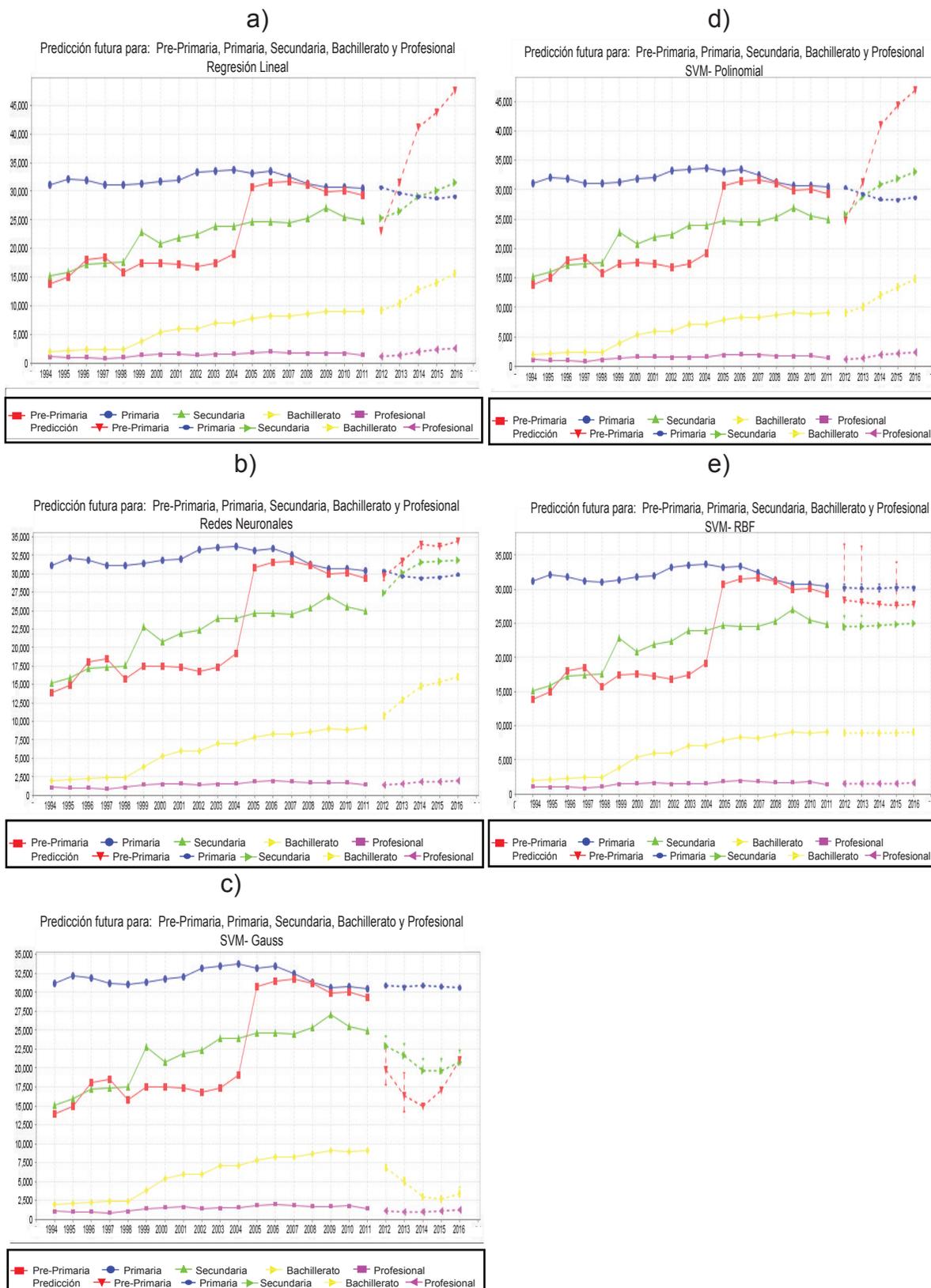


FIGURA 4.10: Resultados obtenidos utilizando diferentes métodos de regresión a) Regresión lineal b) Redes neuronales c) SVM método de Gauss d) SVM-Polinomial e) SVM-RBF

## 4.7. Cuautitlán de Romero Rubio

Cuautitlán de Romero Rubio presenta según la Figura 4.9 un comportamiento similar al de Chimalhuacán en algunas de sus variables como en Pre Primaria. Primaria presenta en contraste con otros municipios estudiados disminución en el número de egresados su máximo se presenta en el 1996 y su mínimo en 2009 con una continua caída, salvo los dos últimos años del periodo de estudio con datos reales. Pre Primaria por el contrario muestra una cresta en 1996, seguida de un valle que se prolonga hasta 2002 para enseguida mostrar un rápido y grande incremento en sus cifras, aunque no es su máximo, para 2006 ya ha alcanzado los niveles de primaria, para 2007 presenta el máximo de toda la gráfica, el periodo siguiente es de estancamiento hasta el 2011. Secundaria tiene un crecimiento con dos saltos prominentes en 1999 y 2010, el máximo esta en 2010 su crecimiento se ve estancado de 2006 a 2011. Bachillerato tiene crecimiento exponencial del inicio a 1999 una recesión y enseguida en 2001 continua el crecimiento relativamente uniforme hasta el final. Profesional-Técnico tiene en contraste con otros municipios un crecimiento más apreciable no uniforme en su totalidad, pero en la tendencia general positivo.

En cuanto a las predicciones, RL y SVM Pol (Figura 4.9 a) y d)) son de nuevo muy parecidas en general, SVM Gauss (Figura 4.9 c)) pareciera tener un comportamiento también similar a los dos modelos antes descritos, solamente que algunas de las líneas están ya sea desfasadas verticalmente, una magnitud o escala reducida en el caso de Pre Primaria, o con una pendiente diferente. En esta ocasión, no antes visto antes en los demás municipios SVM RBF y RNA parecieran tener más similitudes a simple vista, pero varias características hacen que esto no sea verdad, Primaria con pendiente negativa converge en el último año con Pre Primaria con pendiente positiva, formando un triángulo isósceles acostado en el modelo de RNA, el modelo de RBF tiene la misma figura de triángulo, solo que con Pre Primaria con al pendiente negativa y Primaria con la pendiente Positiva juntándose en el último año. En Profesional-Técnico el mejor escenario lo muestra RNA, seguido por SVM Pol. que tienen pendientes positivas más pronunciadas que en los demás modelos.

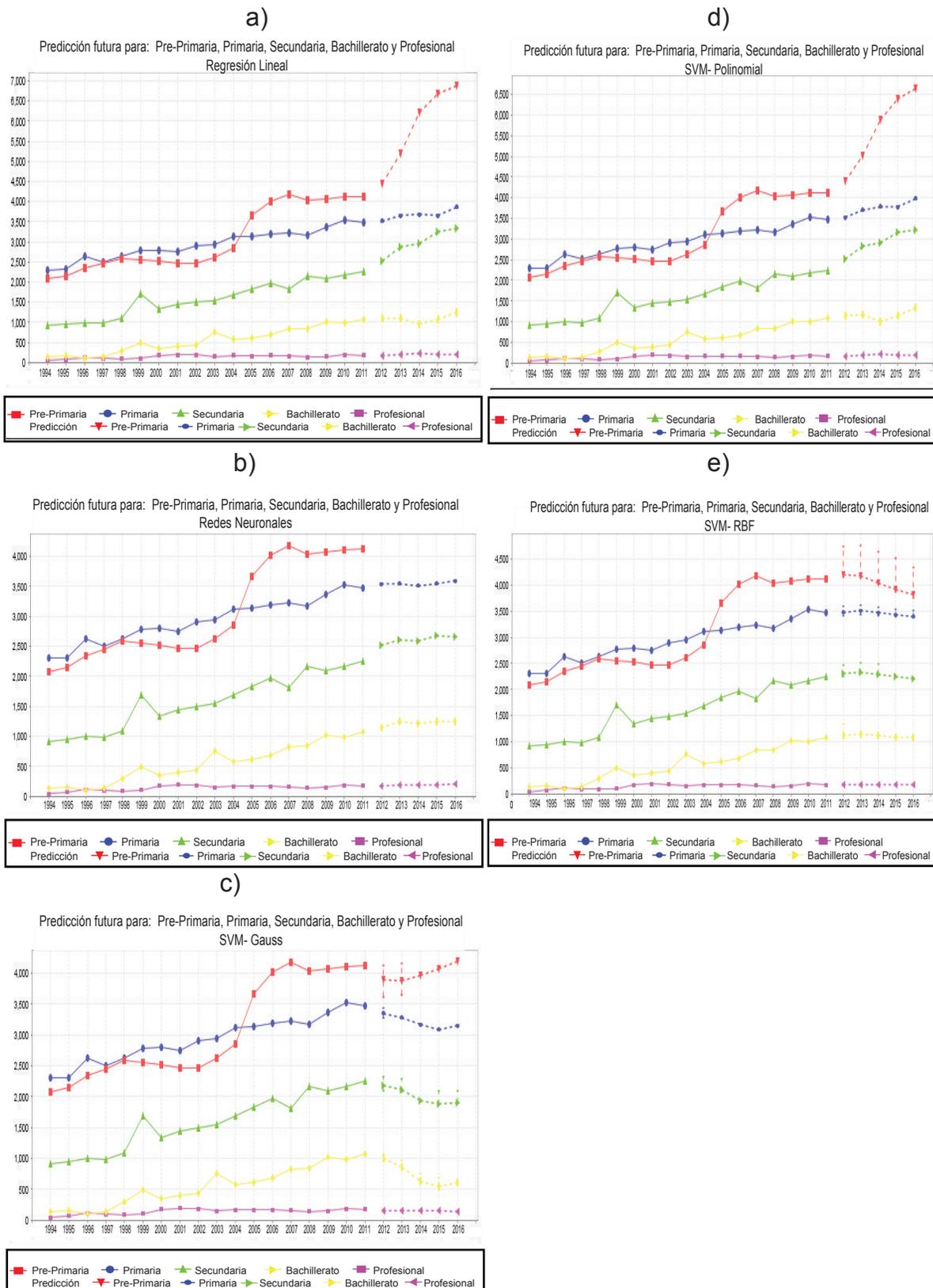


FIGURA 4.11: Resultados obtenidos utilizando diferentes métodos de regresión a) Regresión lineal b) Redes neuronales c) SVM método de Gauss d) SVM-Polinomial e) SVM-RBF

## 4.8. Ecatepec

El municipio de Ecatepec tiene un comportamiento similar al de Cuautitlán (Figura 4.10), en la mayoría de sus variables, solo que, con una magnitud mayor en respuesta a la mayor población de dicho municipio. Primaria muestra un comportamiento sin crecimiento a largo plazo, por el contrario, a partir de 2009, los últimos tres años del estudio disminuyó aún más la cantidad de egresados de los que había en 1994. Pre-primaria muestra un salto entre 2004 y 2005, similar al comportamiento de Cuautitlán. Secundaria tiene dos picos en el 1999 y en el 2009, como en el mencionado municipio, Bachillerato y Profesional-Técnico también.

RL y SVM Pol. (Figura 4.10 a) y d)) nuevamente se parecen mucho, las intersecciones de sus líneas son muy similares, solo en la de Primaria con Secundaria hay una mayor distancia entre ambas después de la intersección. SVM RBF por su lado tiene tendencias en sus predicciones muy lineales, sin grandes variaciones, Pre Primaria es la que tiene una mayor pendiente negativa. SVM Gauss tiene comportamientos casi lineales y sin altibajos en Primaria y Profesional-Técnico, el resto de variables tiene un comportamiento de cuchara que baja y vuelve a subir, solo Bachillerato muestra un descenso del que no se recupera.

## 4.9. Huixquilucan

Huixquilucan, Figura 4.11, muestra una tendencia positiva en casi todos los niveles de educación, Pre Primaria tiene salto después del 2005, son referencia del preescolar obligatorio comenzado por esos años. Primaria tienen uniformidad en su crecimiento. Secundaria presenta uniformidad también, a excepción de un pico y una caída en 1999 y 2007 respectivamente. Bachillerato también posee un pico en 1999 y en 2003. Profesional-Técnico no muestra nada importante en su movimiento, después de 2001, comienza a decaer pululantemente, 2001 es su punto máximo, y después su punto más bajo es 2008.

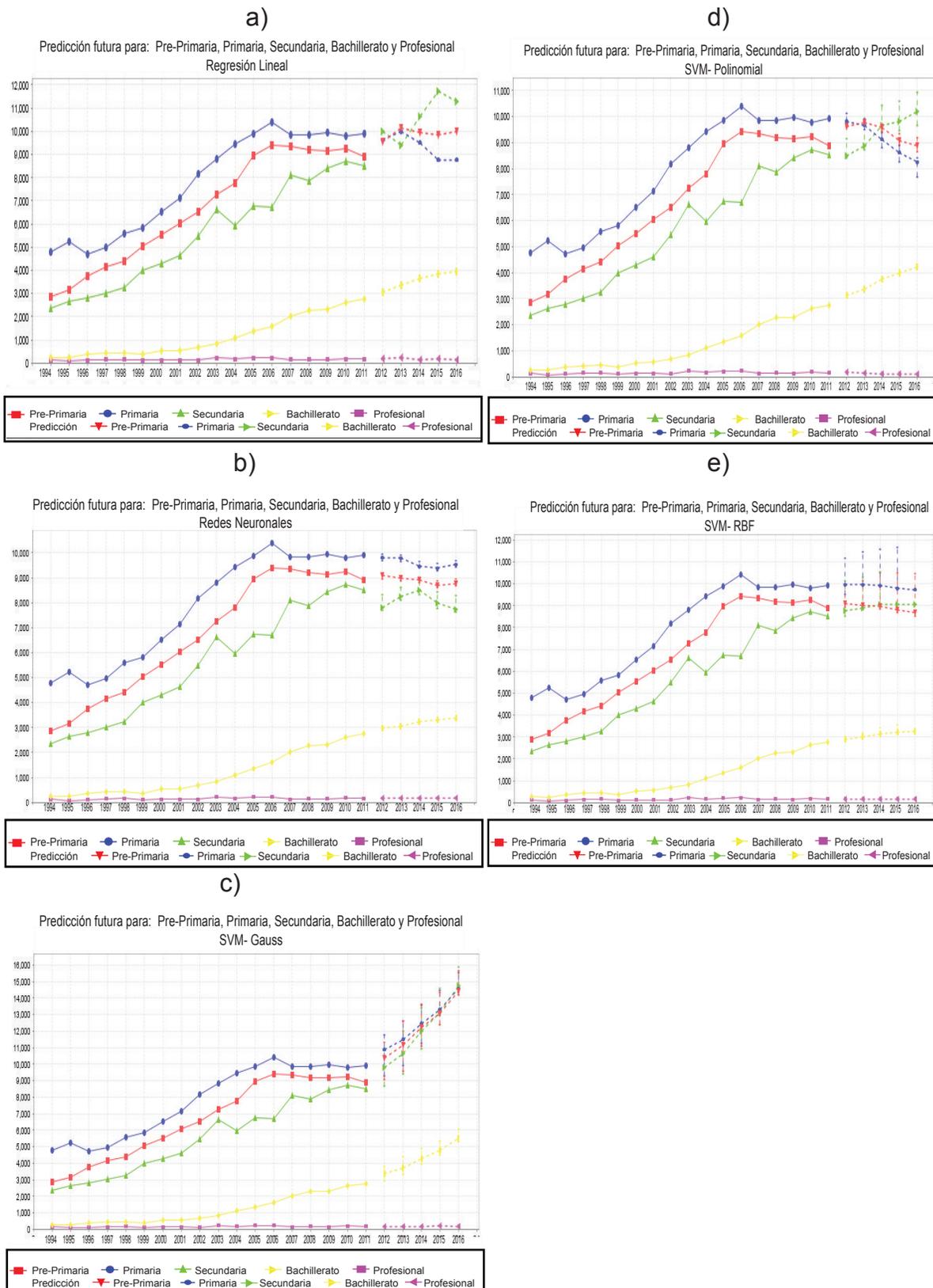


FIGURA 4.12: Resultados obtenidos utilizando diferentes métodos de regresión a) Regresión lineal b) Redes neuronales c) SVM método de Gauss d) SVM-Polinomial e) SVM-RBF

Las predicciones tienen diferentes comportamientos, SVM Gauss (Figura 4.11 c)) nos dice que habrá descenso en las cifras de todas las variables exceptuando Profesional-Técnico, y posteriormente vendrá un incremento, en diferente año para los diferentes niveles educativos según la gráfica. Profesional-Técnico permanece casi recto con una pendiente ligeramente negativa. RL y SVM Pol. (Figura 4.11 a) y d)) tienen comportamientos similares Pre Primaria es la variable con mejores cifras a futuro con otro salto entre 2012 y 2014 incrementando un 70 % respecto a su mejor punto. El escenario de Primaria en positivo pero su crecimiento es uniforme, guarda una pendiente muy parecida a la misma con la que viene creciendo durante el periodo de datos reales; Para Bachillerato nos dice que va a haber un descenso y enseguida una recuperación una pequeña cuchara. Profesional-Técnico menciona que estará con cifras similares en los siguientes cinco años tal vez con un poco más de egresados en 2014. SVM RBF nos pronostica que habrá en los siguientes dos años aproximadamente, cifras similares a la del último año con datos reales (2011), pero en seguida comenzará a disminuir en los cuatro primeros niveles educativos menos pronunciado en primaria y más pronunciado en Pre Primaria. RNA nos dice que la mayoría de datos futuros serán similares en los años posteriores, con ligeras variaciones, en esta ocasión presenta pronósticos casi lineales.

## 4.10. Ixtapaluca

Ixtapaluca muestra un comportamiento en sus datos (Figura 4.12) un crecimiento ajustable a una curva logarítmica en Pre Primaria, Primaria y Secundaria, y exponencial en Bachillerato, en Profesional-Técnico es muy parecido a una línea recta. Primaria supera en todos los años al resto de variables.

Ixtapaluca muestra en el modelo RNA (Figura 4.12 b)) predicciones tendencias no tan positivas, en Primaria es una tendencia negativa con forma de “cuchara” pero con el último dato menor al primero del periodo de predicción. Pre Primaria también es negativa. Secundaria tiene una tendencia al alza y posteriormente desciende nuevamente teniendo un máximo local de 2012 a 2016, sin embargo, el máximo está al nivel del último dato real. Bachillerato es la variable con mejores expectativas en una línea semirrecta con pendiente positiva. Profesional no muestra sino una continua línea recta respecto al resto del historial de datos reales. SVM RBF por otra parte muestra lianas semirrectas con pendiente positiva en Secundaria y Bachillerato, y pendiente negativa en Primaria y Pre Primaria. SVM Gauss (Figura 4.12 c)) tiene un escenario positivo para todas las variables menos Profesional-Técnico, los cuatro primeros niveles educativos

tienen una figura con tendencia a una curva exponencial, Profesional-Técnico tiene una tendencia recta y aun negativa en el último año del pronóstico. RL y SVM Pol. Esta ocasión tienen menos pero no nulas coincidencias, Primaria y Pre Primaria son parecidas en forma, Secundaria es irregular, pero con una tendencia positiva en ambas.

#### 4.11. La Paz

La Paz muestra un comportamiento (Figura 4.13) en Primaria oscilante sin progreso destacable al largo plazo, Pre Primario por el contrario ha mostrado un incremento importante en sus cifras, ha crecido por encima del doble en el periodo de 17 años, con saltos periódicos destacando el del año 2005.

En las predicciones SVM Pol y RL son muy parecidas, en esta ocasión también se ha unido a sus predicciones RNA, el modelo de Redes Neuronales (Figura 4.13 b)) tiene un comportamiento similar el de los dos modelos antes mencionados, con pequeñas diferencias en magnitudes o pendientes, sin embargo, ha resultado muy parecida. SVM RBF muestra tendencias de pronóstico semilineales, salvo Pre Primaria con una inclinación cóncava de magnitud negativa.

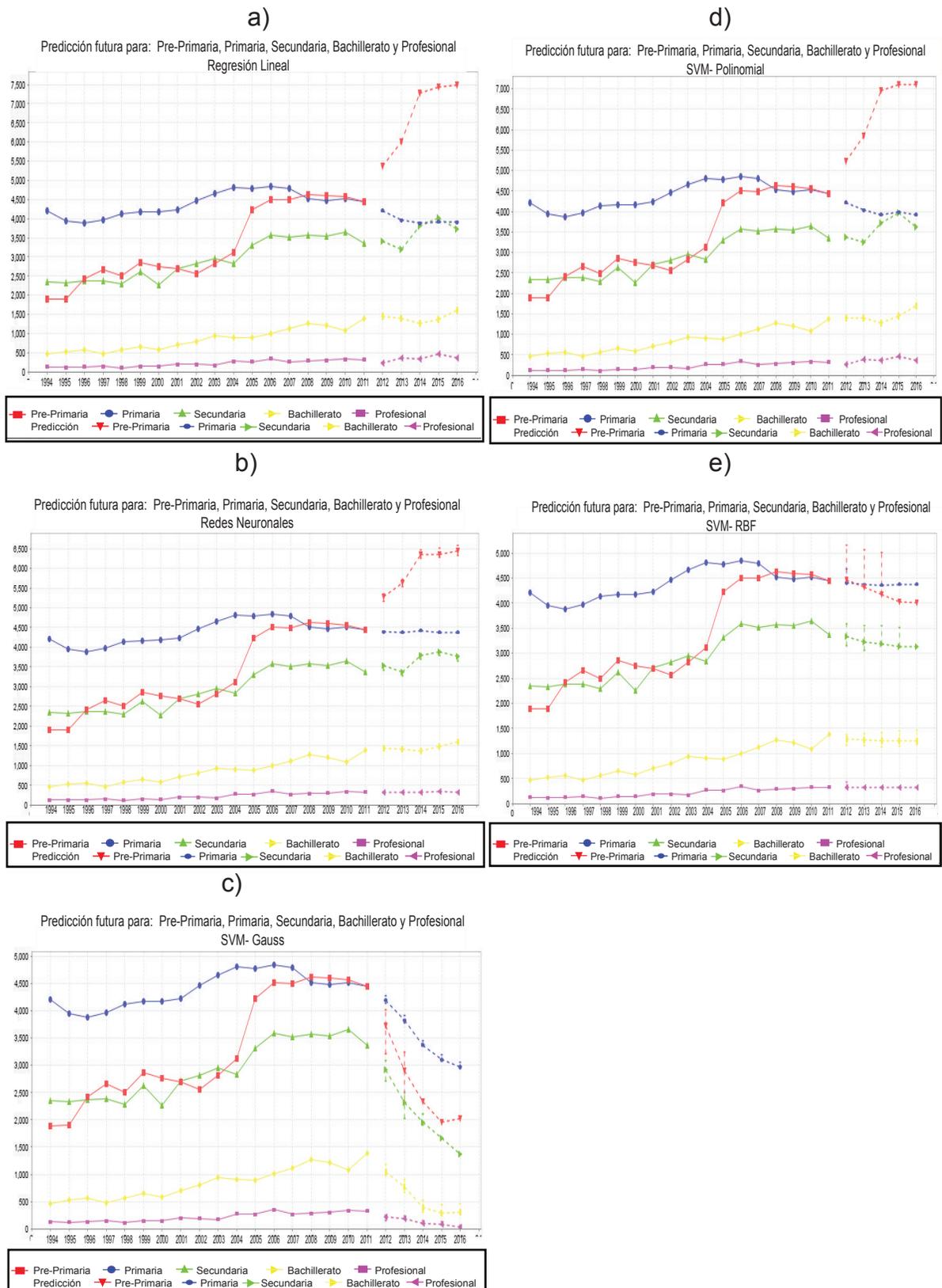


FIGURA 4.13: Resultados obtenidos utilizando diferentes métodos de regresion a) Regresión linealr b) Redes neuronales c) SVM método de Gauss d) SVM-Polinomial e) SVM-RBF

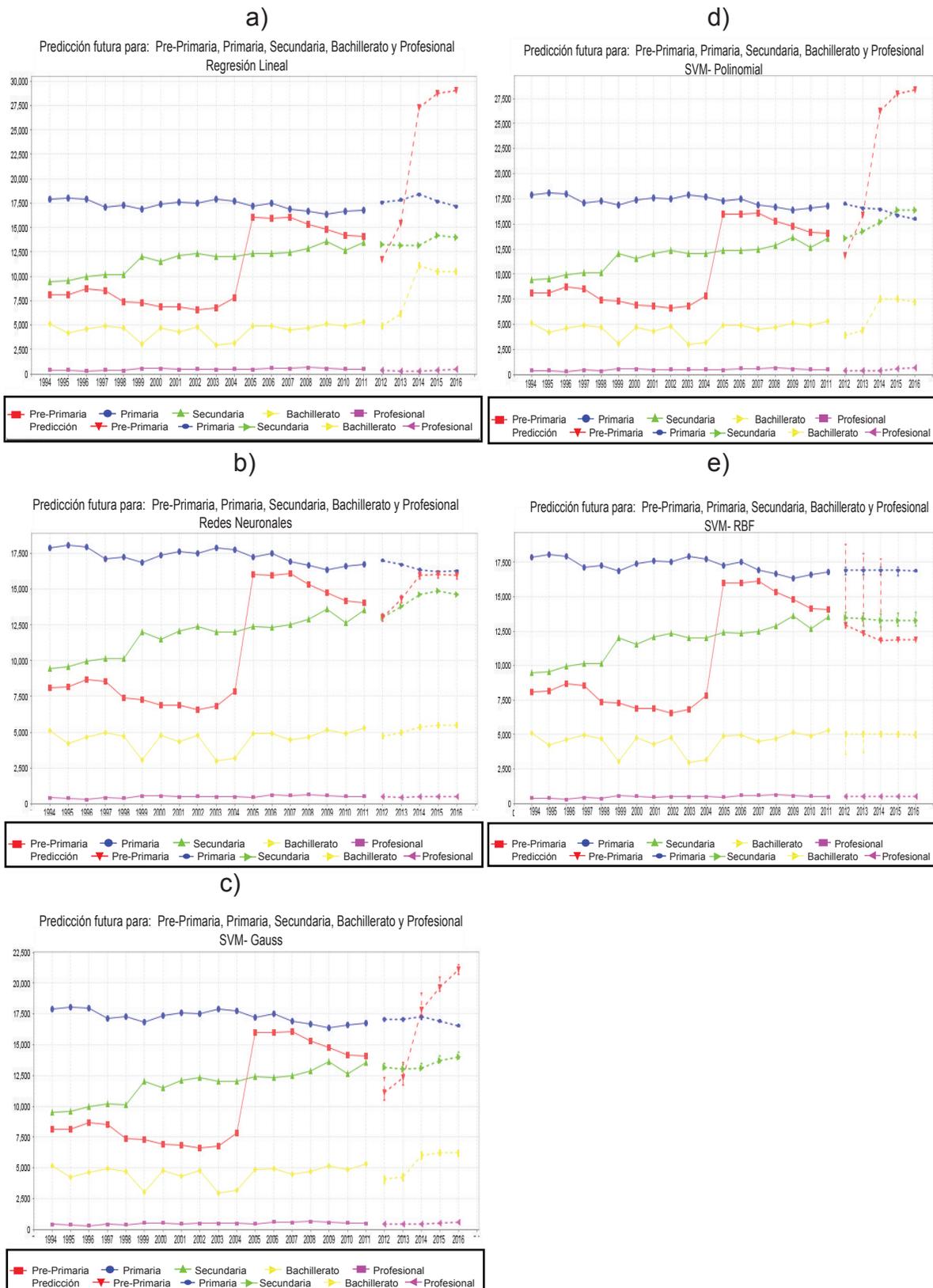


FIGURA 4.14: Resultados obtenidos utilizando diferentes métodos de regresión a) Regresión lineal b) Redes neuronales c) SVM método de Gauss d) SVM-Polinomial e) SVM-RBF

## 4.12. Naucalpan

La Figura 4.14 muestra el desarrollo de los egresados de Naucalpan; Al largo plazo tiene una caída paulatina en Primaria, pequeñas variaciones, pero en general negativa, Pre Primaria ha resultado en datos positivos, sobre todo en 2005 donde casi se duplica la cantidad de egresados del nivel académico, su tendencia parece ser menguante salvo el año 2005 y algunos años en que sube poco para enseguida seguir disminuyendo. Secundaria tiene una actitud creciente en todo el periodo de estudio. Bachillerato he mostrado un estancamiento en sus cifras, con años peores que otros, pero no logra salir del mismo nivel de egresados. Profesional-Técnico es casi lineal.

La Figura 4.14 c) a) y d) muestra los datos de las proyecciones de SVM Gauss. RL y SVM Pol. tienen un comportamiento de salto en Pre Primaria, el salto es mayor al que presenta el modelo Gauss sin embargo es importante destacar que los tres modelos lo presentan de una manera tan clara; Gauss comparte con RL los comportamientos de Pre Primaria y Primara, comparte con SVM Pol. figuras muy similares en Bachillerato, desde luego en Profesional-Técnico, ya que es casi lineal. El modelo SVM RBF tiene comportamientos semilineales en todas las variables, salvo en Pre Primaria que tiene un declive los tres primeros años de predicción y los otros dos vuelven a estabilizar la pendiente a cero. EL modelo ARN es el modelo que tiene Primaria y Secundaria con la orientación de puntos de inflexión de las funciones de tendencia contrarios a los que presenten los tres primeros modelos descritos, RL, SVM Gauss y SVM Pol., Pre Primaria muestra un incremento mas no tan enorme como el de los primeros modelos.

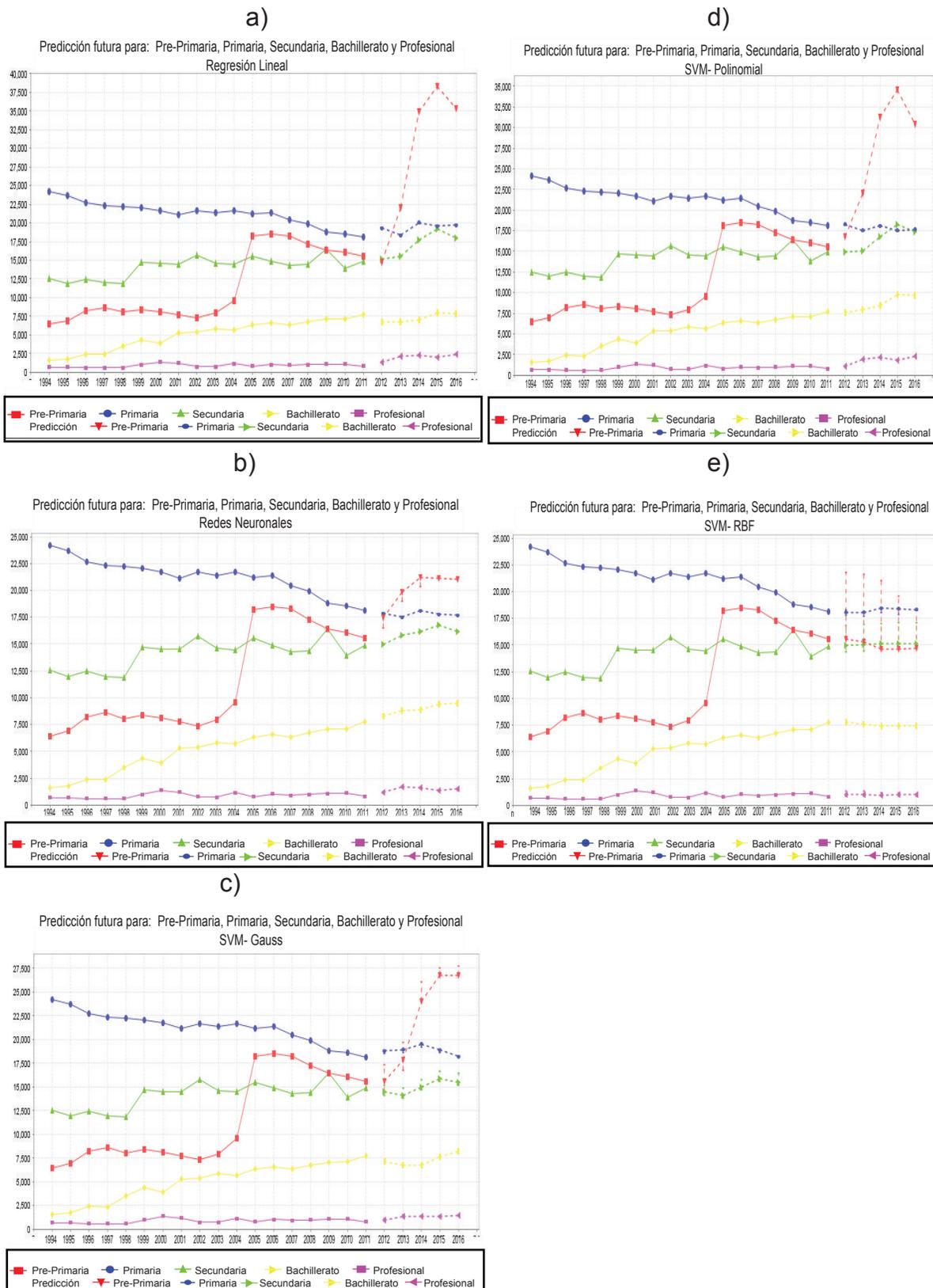


FIGURA 4.15: Resultados obtenidos utilizando diferentes métodos de regresión a) Regresión lineal b) Redes neuronales c) SVM método de Gauss d) SVM-Polinomial e) SVM-RBF

## 4.13. Nezahualcóyotl

En Nezahualcóyotl, se aprecia en la Figura 4.15 un comportamiento negativo en Primaria, casi no hay puntos de aumento importantes, Pre Primaria no alcanza a Primaria, pero supera el doble de cifras del inicio al fin del periodo de estudio, Secundaria tiene incremento de inicio a fin, mas con muchas variaciones. Bachillerato es más uniforme y creciente casi todo el periodo. Profesional-Técnico e casi lineal, pero con unas pequeñas variaciones.

La predicción de Nezahualcóyotl por el modelo de RNA (Figura 4.15 b)) es distinta el resto de las predicciones; RL y SVM Pol. son muy parecidas en su predicción, muestran en Pre Primaria como le vemos en Figura 4.15 un alto pico en Pre Primaria, que, aunque con una figura un poco distinta en la parte más alta, Gauss también presenta un salto de magnitud parecida a estos dos modelos; Solo el comportamiento de Gauss de Primaria coincide en forma y vértices con RNA. SVM RBF tiene tendencias más lineales que el resto de modelos con ligeros cambios de pendiente tras algunos años peor si más variaciones que se parezcan a las de algún otro modelo.

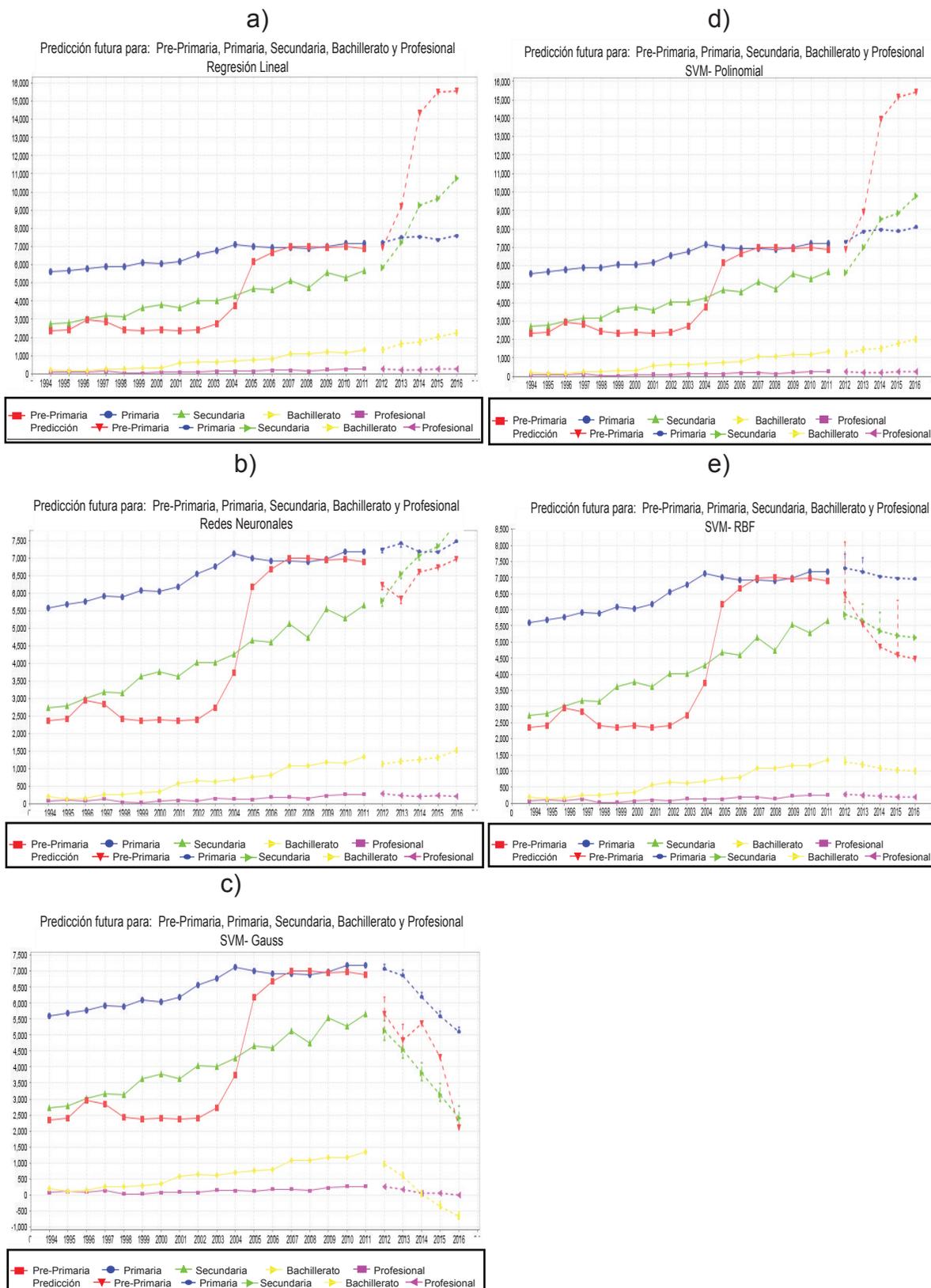


FIGURA 4.16: Resultados obtenidos utilizando diferentes métodos de regresion a) Regresión lineal b) Redes neuronales c) SVM método de Gauss d) SVM-Polinomial e) SVM-RBF

## 4.14. Nicolás Romero

La Figura 4.16 nos muestra el comportar de las variables reales, Primaria muestra una tendencia positiva hasta el 2004 de allí en adelante una recesión donde hay recuperación hasta el 2010, Pre Primaria muestra la misma curva de salto del 2005 que continúa creciendo hasta 2007 de allí cae ligera y paulatinamente hasta 2011. Secundaria es creciente de inicio a fin de esta duplicando, con un patrón de oscilamiento. Bachillerato en comparación con el resto de variables es bastante bajo, también en comparación con la diferencia entre Bachillerato y sus niveles precedentes en otros municipios. Por ultimo Profesional-Técnico en muy bajo pero creciente.

Las predicciones de Nicolás Romero, tienen características similares en SVM Pol y RL (Figura 4.16 d) y a)); SVM Gauss y SVM RBF (Figura 4.16 c) y e)) tienen tendencias muy negativas, Gauss presenta líneas semirectas de pendiente negativa en Primaria, Secundaria y Bachillerato, Profesional-Técnico también es pendiente negativa pero mucho menos inclinada que las otras tres variables. Solo Pr Primaria tiene irregularidades en su tendencia que al final sigue negativa en su pendiente. Para concluir, SVM RBF tiene líneas semirectas de pendiente negativa en todas las variables excepto Pre Primaria, que es una curva de inclinación negativa.

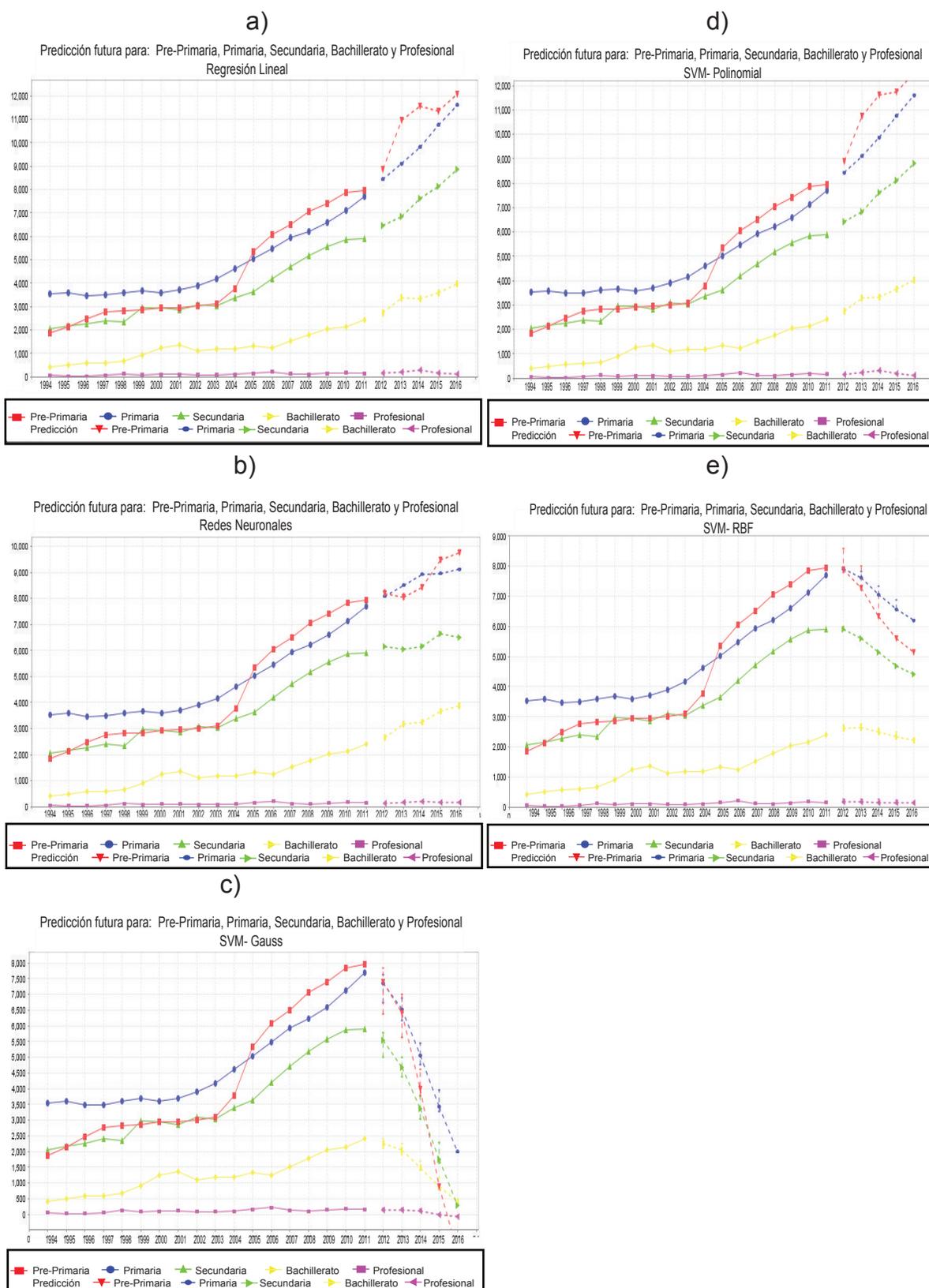


FIGURA 4.17: Resultados obtenidos utilizando diferentes métodos de regresión a) Regresión lineal b) Redes neuronales c) SVM método de Gauss d) SVM-Polinomial e) SVM-RBF

## 4.15. Tecamac

Según la Figura 4.17 las líneas de crecimiento de las variables en Tecámác son bastante suaves comparadas con el resto de municipios estudiados, casi sin picos que rompan la continuidad en los datos, exceptuando Profesional-Técnico todas las variables son crecientes, Pre Primaria destaca por su rápido aumento en 2005. Se caracterizan por tener una figura de “S” invertida al crecer.

La Figura 4.17 c) nos muestra un escenario del modelo SVM Gauss donde apreciamos caídas muy notables que juntamente con los datos reales forman la figura de una campana de Gauss. SVM RBF (Figura 4.17 e)) por otro lado tiene también resultados adversos, pero no tan negativos como los de Gauss. SVM Pol. y RL guardan similitudes importantes. ARN por su lado tiene una tendencia relativamente uniforme para Primaria, Positiva para Pre Primaria, positiva para Secundaria y Bachillerato, igualmente Profesional-Técnico, pero este último con menos velocidad de crecimiento.

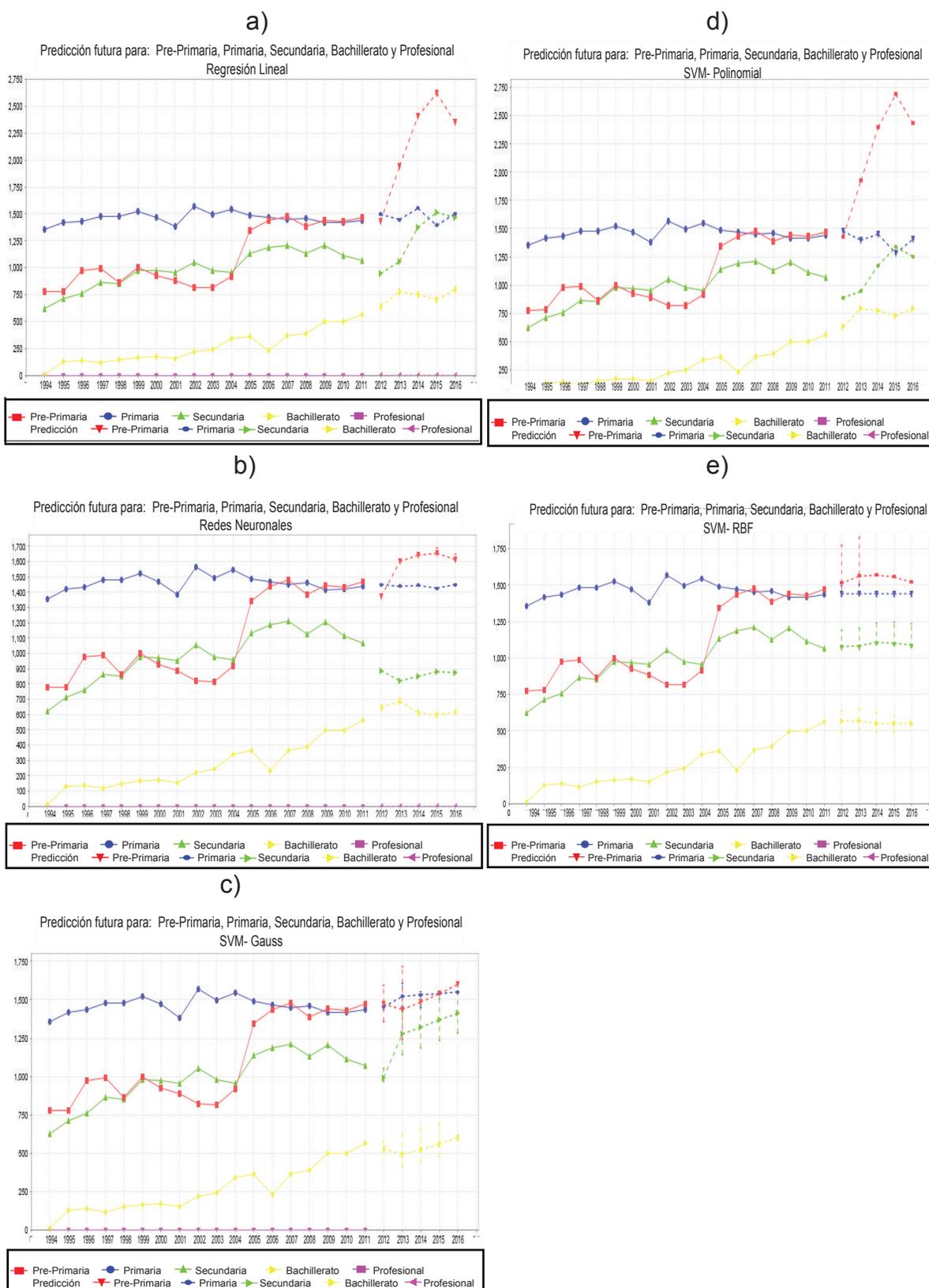


FIGURA 4.18: Resultados obtenidos utilizando diferentes métodos de regresión a) Regresión lineal b) Redes neuronales c) SVM método de Gauss d) SVM-Polinomial e) SVM-RBF

## 4.16. Tepetzotlán

Tepetzotlán con los datos de la Figura 4.18 apreciamos Primaria en un periodo de altibajos y poco crecimiento, Pre Primaria también muestra esa clase de variaciones, pero con el salto del 2005, Secundaria es creciente con muchas variaciones. Bachillerato igualmente es creciente de forma irregular. Profesional-Técnico es también bajo respecto a las demás variables.

La Figura 4.18 c) presentada es del modelo SVM Gauss, sus predicciones son de forma irregular en comparación con SVM RBF por cuanto dicho modelo guarda resultados muy lineales a excepción de Pre Primaria que es convexo. Gauss tiene un salto en Secundaria de donde ahora crece de forma gradual. Pre Primaria tiene una caída tras la cual inmediatamente comienza a subir a Primaria tiene la misma figura, pero con posición invertida. Bachillerato es cóncava con pérdidas seguidas de ganancias. RNA (Figura 4.18 b)) también es distinto, Pre Primaria con una curva log del 2012 al 2015 para descender inmediatamente un poco, Primaria está estancada sin crecimiento ni recesiones importantes, Secundaria tiene un escenario menos favorable de estancamiento, peor por debajo de los datos de 2011. Bachillerato es oscilante de arriba abajo.

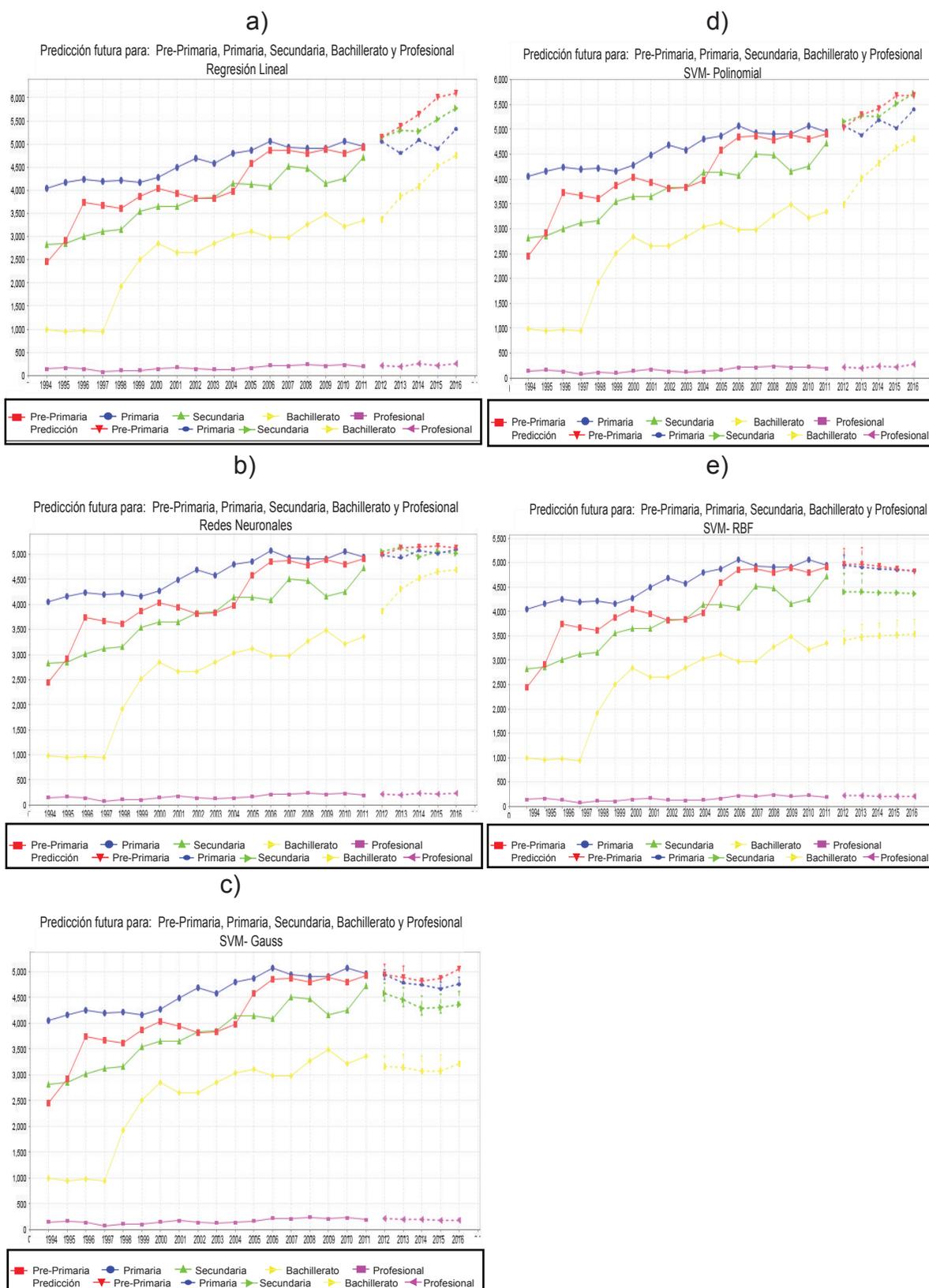


FIGURA 4.19: Resultados obtenidos utilizando diferentes métodos de regresión a) Regresión lineal b) Redes neuronales c) SVM método de Gauss d) SVM-Polinomial e) SVM-RBF

## 4.17. Texcoco

La Figura 4.19 muestra una tendencia positiva en Primaria, Pre Primaria y Secundaria, muy alta en Bachillerato y aun ligeramente en Profesional-Técnico. Bachillerato tiene un salto del 1997 al 2000 creciendo de inicio a fin más del triple de egresados. El comportamiento de las variables es fluctuante.

Texcoco tiene similitudes grandes en los modelos SVM Pol. y RL (Figura 4.19 d) y a)). El modelo de Gauss nos muestra que las tendencias son con forma de “cuchara” todas ellas (Figura 4.19 b)). SVM RBF dice que las tendencias son semilineales y casi horizontales (Figura 4.19 e)). RNA nos dice que los pronósticos para los tres niveles más básicos están muy juntos entre si con vista a ganar, Bachillerato tiene un comportamiento logarítmico.

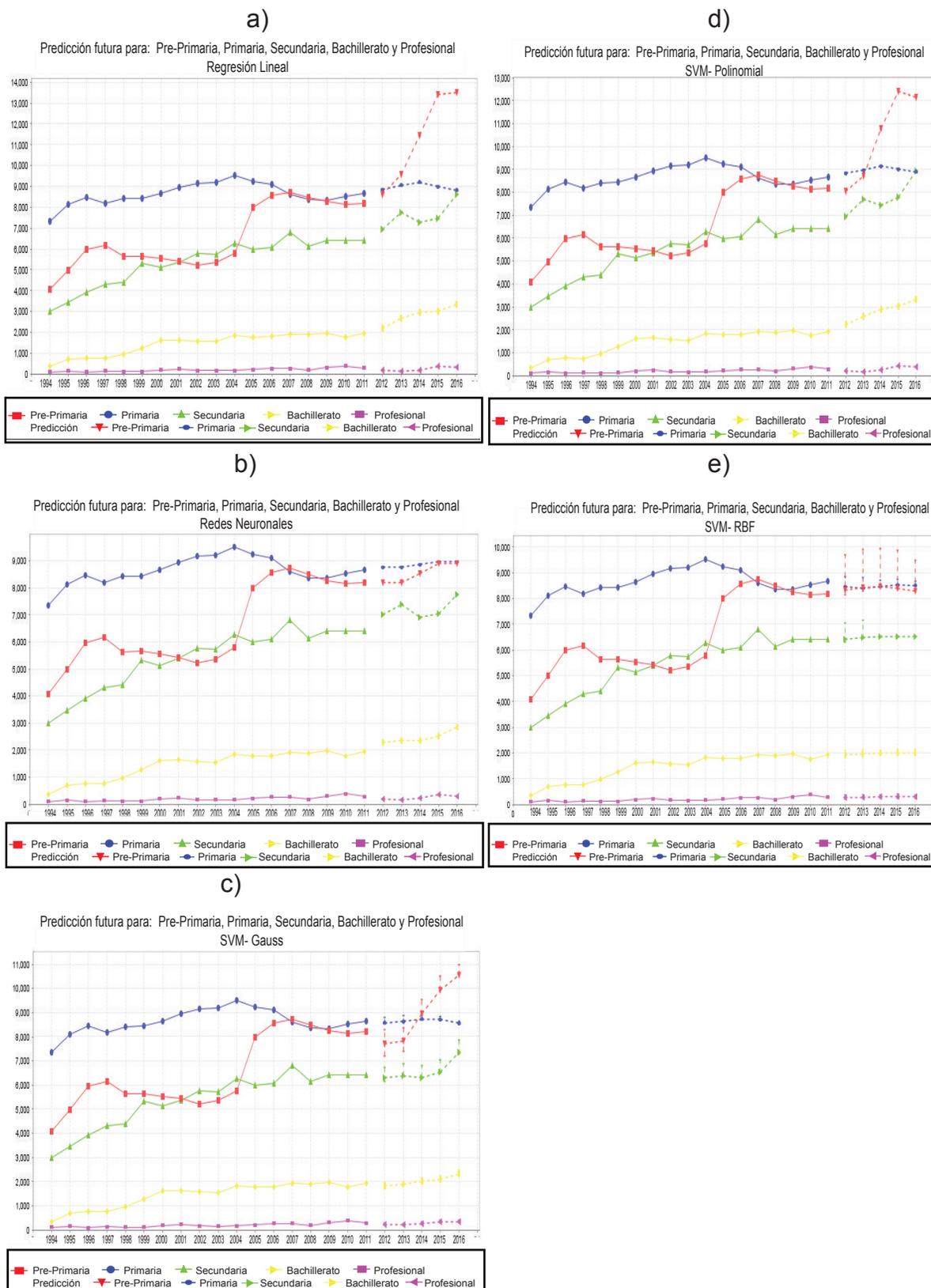


FIGURA 4.20: Resultados obtenidos utilizando diferentes métodos de regresión a) Regresión lineal b) Redes neuronales c) SVM método de Gauss d) SVM-Polinomial e) SVM-RBF

## 4.18. Tlalnepantla

La Figura 4.20 muestra un crecimiento lento en Primaria, por los altibajos hacen que tenga variaciones, Pre Primaria también tiene su crecimiento mayor en 2005. Secundaria tiene crecimiento positivo. Bachillerato, comparado con Secundaria. Profesional-Técnico igual muestra tendencia de crecimiento mucho más ligero que el resto de las variables.

Tlalnepantla muestra en la Figura 4.20 d) el comportamiento del modelo SVM Pol. que es similar al RL. El modelo de SVM Gauss tiene todas las variables exceptuando Profesional-Técnico tienen forma cóncava de “cuchara” donde bajan y vuelven a subir enseguida. El modelo SVM RBF guarda su patrón de líneas semirectas con pendiente cercana a cero. Por último, el modelo RNA tiene un escenario ligeramente positivo para Primaria, Pre Primaria forma junto con los datos reales un columpio del que se recupera en 2015 y 2016. Secundaria muestra altibajo como lo hizo durante su historial de datos, Bachillerato es exponencial y Profesional-Técnico en forma de “S”.

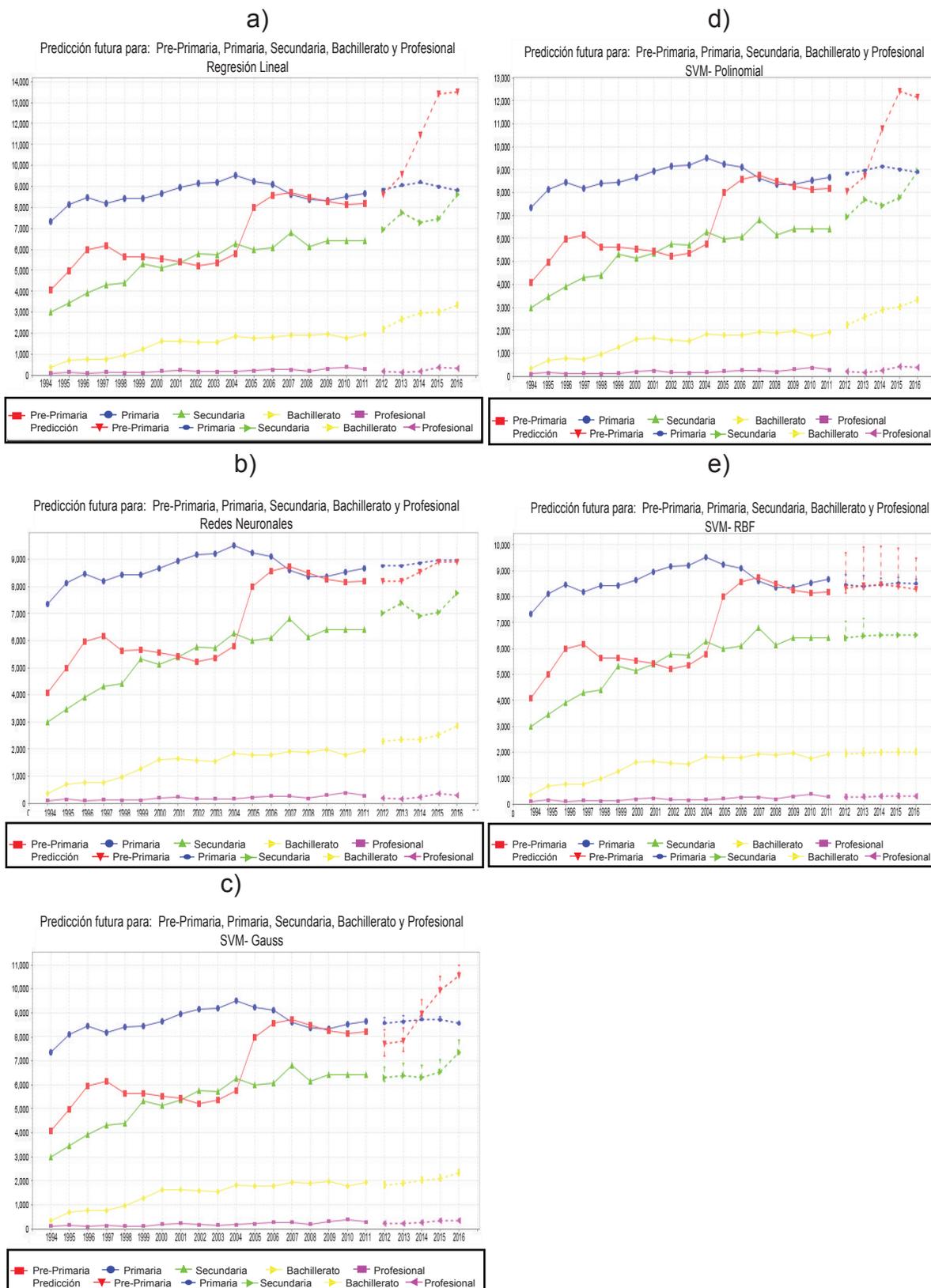


FIGURA 4.21: Resultados obtenidos utilizando diferentes métodos de regresión a) Regresión lineal b) Redes neuronales c) SVM método de Gauss d) SVM-Polinomial e) SVM-RBF

## 4.19. Tultitlan

La Figura 4.21 muestra un crecimiento lento en Primaria, por los altibajos hacen que tenga variaciones, Pre Primaria también tiene su crecimiento mayor en 2005. Secundaria tiene crecimiento positivo. Bachillerato, comparado con Secundaria. Profesional-Técnico igual muestra tendencia de crecimiento mucho más ligero que el resto de las variables.

Tultitlan muestra en la Figura 4.21 e) el comportamiento del modelo SVM RBF que guarda su patrón de líneas semirrectas con pendiente cercana a cero. El modelo de SVM Gauss tiene todas las variables exceptuando Profesional-Técnico tienen forma cóncava de “cuchara” donde bajan y vuelven a subir enseguida. Los modelos SVM Pol. y RL son muy parecidos. El modelo RNA tiene en Primaria un ligero incremento, Pre Primaria forma junto con los datos reales una “cuchara” como en el pasado municipio. Secundaria muestra el incremento al siguiente año y en seguida hace una curva de “cuchara” hasta terminar el periodo en 2016, Bachillerato es exponencial también y Profesional-Técnico tiene forma de “S”.

## 4.20. Análisis comparativo de los datos Reales VS Predicciones

En esta sección se llevará a cabo un análisis comparativo entre los datos reales y los pronosticados por los diferentes modelos, para poder así medir su ajuste a los datos reales, el objetivo es lograr ver cuál es el mejor modelo que mejor se ajusta alzo datos reales, se presentaran datos e información de algunas métricas indicadoras basadas en error para realizar las comparaciones. En esta sección para este estudio, únicamente se presenta uno de los municipios estudiados, Tlalnepantla, para hacer el análisis completo de su precisión, se tomaron en cuenta los últimos cinco años de donde se tienen datos reales, 2007 a 2011, para estudiar. Se han de aplicar estudios a todos los niveles de escolaridad, Pre-Primaria, Primaria, Secundaria, Bachillerato y Profesional-Técnico, en Tlalnepantla, las métricas también se aplican todas en los niveles mencionados. Se ha determinado que por causa del espacio únicamente se presentara el municipio de Tlalnepantla en el estudio.

El municipio de Tlalnepantla se estudió previamente en la sección anterior, para descubrir que el comportamiento en sus variables es difiere según el nivel académico, Primaria tiene una predominancia general en comparación con el resto de las variables,

CUADRO 4.1: Valores reales conjunto de datos Tlanepantla

Año	Pre-Primaria	Primaria	Secundaria	Preparatoria	Universidad
1994	4075	7345	2995	359	107
1995	4986	8114	3460	703	150
1996	5973	8455	3916	770	103
1997	6154	8191	4307	752	142
1998	5631	8412	4396	968	115
1999	5642	8436	5322	1257	122
2000	5542	8647	5128	1609	200
2001	5428	8942	5381	1641	232
2002	5207	9164	5774	1576	174
2003	5350	9200	5727	1549	160
2004	5778	9520	6270	1835	173
2005	7987	9236	5990	1782	217
2006	8554	9092	6085	1795	271
2007	8734	8603	6796	1920	271
2008	8484	8355	6136	1883	190
2009	8264	8343	6411	1974	302
2010	8143	8514	6411	1770	382
2011	8195	8661	6416	1932	284

su comportamiento no tiene grandes variaciones como otras variables, es uniforme, lo que permite que Pre Primaria lo alcance y súper en el 2007 y 2008 aunque posteriormente vuelve a ser menor que Primaria, esta variable tiene una tendencia de crecimiento de salto por temporadas. Por otro lado, secundaria, tiene un crecimiento más uniforme que sus predecesoras, no así Bachillerato o Preparatoria que es los mismo y Profesional que superan a Secundaria en linearidad en cifras mucho menores. Si consideramos los últimos cinco años de datos reales tendemos que respecto al resto de la línea histórica hay un comportamiento muy estable sin tantos picos o saltos como en años previos, aun así, podemos estudiar el ajuste que tienen con los datos predichos. Primeramente, se realizarán los estudios basándonos en las gráficas donde vemos las líneas de comportamiento.

Para Pre Primaria tenemos una línea en los datos reales decreciente, salvo en el último año que tiene un pequeño incremento, podemos ver que las líneas de SVM-RBF o SVM- Gauss esta distantes de la línea de los datos verdaderos, SVM-Pol y RNA parecen ajustarse mas a los datos correctos, RL por otro lado parece ajustarse también, aunque no en la medida de las dos anteriores. La línea de Pre Primaria a mostrado saltos por tiempo, en los que muestra brincos seguidos de un periodo prolongado de decaimiento progresivo hasta el próximo salto, en general su tendencia es creciente.

CUADRO 4.2: Comparativa de datos reales y resultados utilizando las diferentes técnicas de regresión

	Pre-Primaria	Primaria	Secundaria	Preparatoria	Universidad
<b>2007</b>	<b>8734</b>	<b>8603</b>	<b>6796</b>	<b>1920</b>	<b>271</b>
Predicción LR	8274.3624	9145.623	6496.4439	1963.5089	326.8291
Predicción NN	8470.8548	9284.7564	622.6581	1981.7816	282.3975
Predicción SVM-Gauss	7874.7411	9097.6362	6127.5357	1795.917	252.2458
Predicción SVM-Poly	8223.8871	9293.598	6595.3019	1986.0953	290.6163
Predicción SVM-RBF	6582.272	9261.0031	6325.2373	1894.4361	228.3223
<b>2008</b>	<b>8484</b>	<b>8355</b>	<b>6136</b>	<b>1883</b>	<b>190</b>
Predicción LR	7827.6476	9629.3775	7258.5773	2229.6883	337.4857
Predicción NN	8395.4991	9426.694	7034.2615	2121.5841	264.0263
Predicción SVM-Gauss	7368.661	9219.4125	6267.7619	1879.7003	232.1644
Predicción SVM-Poly	7856.2076	9691.5115	7420.1953	2303.1309	277.2129
Predicción SVM-RBF	6600.2542	9359.1134	6506.0211	1982.7944	231.5386
<b>2009</b>	<b>8264</b>	<b>8343</b>	<b>6411</b>	<b>1974</b>	<b>302</b>
Predicción LR	7964.2187	10082.3809	7610.5576	2455.2986	381.4189
Predicción NN	8526.247	9455.3995	6971.9695	2174.6663	277.8734
Predicción SVM-Gauss	7440.2973	9400.6479	6233.3557	1956.1577	238.8833
Predicción SVM-Poly	8097.7716	10075.9299	7801.8963	2601.8241	318.208
Predicción SVM-RBF	6868.9945	9403.8208	6539.1933	2014.1996	237.5156
<b>2010</b>	<b>8143</b>	<b>8514</b>	<b>6411</b>	<b>1770</b>	<b>382</b>
Predicción LR	8532.355	10396.2544	8168.0909	2603.4041	383.782
Predicción NN	8875.289	9526.0732	6991.1107	2215.5705	292.8031
Predicción SVM-Gauss	8199.0826	9530.6284	6281.4045	1941.1917	258.0268
Predicción SVM-Poly	8885.3831	10460.1211	8124.1565	2790.1607	365.5074
Predicción SVM-RBF	7146.1592	9326.7049	6466.3563	1987.13	244.5728
<b>2011</b>	<b>8195</b>	<b>8661</b>	<b>6416</b>	<b>1932</b>	<b>284</b>
Predicción LR	9728.3982	10386.782	8345.7	2587.0184	351.5263
Predicción NN	9280.2645	9541.2025	7021.5507	2180.0353	295.2361
Predicción SVM-Gauss	8892.2119	9637.0266	6485.0387	1907.1917	257.6379
Predicción SVM-Poly	9761.2316	10558.113	8294.0668	2724.9393	370.406
Predicción SVM-RBF	7255.3179	9228.9736	6343.7628	1925.0222	242.4726

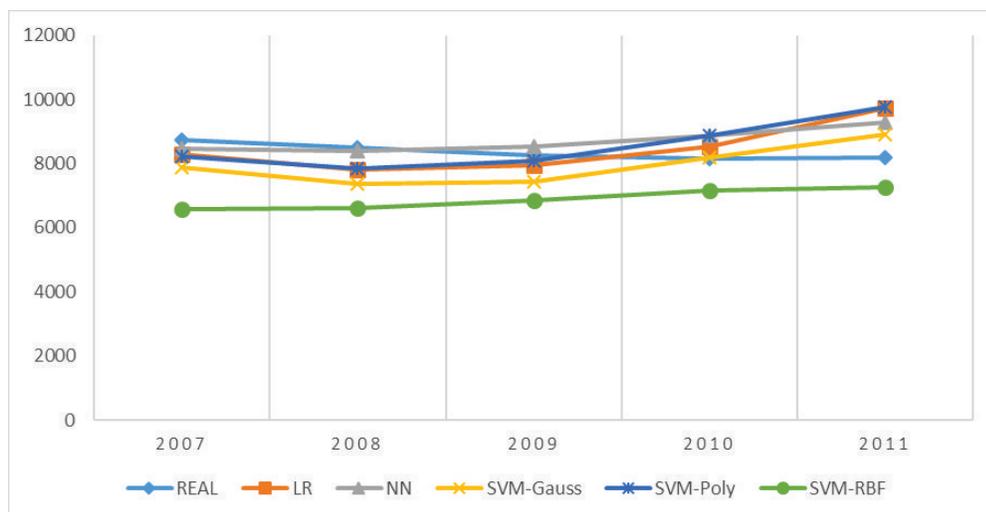


FIGURA 4.22: Predicciones obtenidas para Pre-Primaria

Según las diferentes métricas de error podemos deducir lo siguiente: Mean Absolute Error (MAE) nos dice que el error menor se presenta con el modelo de RNA con 5.81 seguido por SVM-Pol y RL respectivamente, que tienen valores muy similares de poco más de 6 puntos, de allí el salto es enorme a SVM-Gauss y aun mayor hacia SVM-RBF mostrando valores de más de 200 y en el otro de más de 550 puntos respectivamente. En Root Relative Squared Error (RRSE) encontramos el menor valor de error en SVM-Pol, seguido de RNA y RL en ese orden todos con valores en 7 puntos, posteriormente SVM-Gauss con más de 31 puntos y nuevamente al final SVM-RBF con más de 110 puntos. Relative Absolute Error (RAE) favorece a RNA seguido de SVM-Pol con 1.02 y 1.06 respectivamente, RL en 1.11 y otra vez SVM Gauss con valor mayor a 35 y SVM-RBF con mayor a 105 puntos. El indicador Mean Absolute Percentage Error (MAPE) muestra el error en forma de porcentaje respecto al valor real, y nos dice que los menores valores son 10.18 % y 10.27 % para RNA y SVM-Pol respectivamente, sigue 11.5 % que pertenece a RL y para terminar SVM Gauss y SVM-RBF con casi 300 % y 700 % de error en ese mismo orden. Root Mean Squared Error (RMSE) es un indicador grandemente utilizado en la evaluación de predicciones, y está en pro de SVM-Pol con 6.56 seguido de RNA con 6.70 que son los de mejores números, sigue RL con más de 7 puntos y de nuevo saltan los indicadores con SVM-Gauss y SVM-RBF con casi 280 unidades y 1000 unidades para dichos modelos. El último indicador, Mean Squared Error (MSE) que posee la característica de penalizar los errores nos dice que SVM-Pol es el mejor pronóstico con 43.03 que RNA con 44.94, sigue de allí RL con 55 unidades SVM-Gauss con casi 88 mil unidades y SVM-RBF con casi 1 millón de unidades, las cifras por la sensibilidad de la prueba. Analizando estas situaciones encontramos que, de acuerdo a los indicadores tres favorecen a RNA y tres a SVM-Pol.

CUADRO 4.3: Métricas de comparación para el municipio de Tlanepantla (Pre-Primaria)

Métrica	LR	NN	SVM-Gauss	SVM-Poly	SVM-RBF
Mean absolute error	6.8122	5.8154	208.1914	6.1028	558.2244
Root relative squared error	0.7821	0.7486	31.4872	0.7286	112.6818
Relative absolute error	1.1101	1.0274	37.9733	1.0699	106.0009
Mean absolute percentage error	0.115	0.1018	2.9761	0.1027	6.933
Root mean squared error	7.4654	6.7041	279.1539	6.5602	995.7855
Mean squared error	55.7316	44.9451	77926.918	43.0363	991588.7459

Para Primaria se tiene la línea de datos Reales por debajo del resto de líneas, se entiende que todos los modelos de predicción sobrevaloraron los datos en dichos años, se aprecia que el comportamiento de la serie de datos Reales desciende y posteriormente

vuelve a subir durante el periodo de los cinco años; A simple vista SVM-Pol y RL denotan un comportamiento diferente al Real y se aprecia una distancia más amplia a la línea de los mismos que en el resto de las series, SVM-Gauss y SVM-RBF por su lado tampoco parecen ajustarse a la figura de datos verdaderos, el más ajustado podría ser RNA a simple vista.

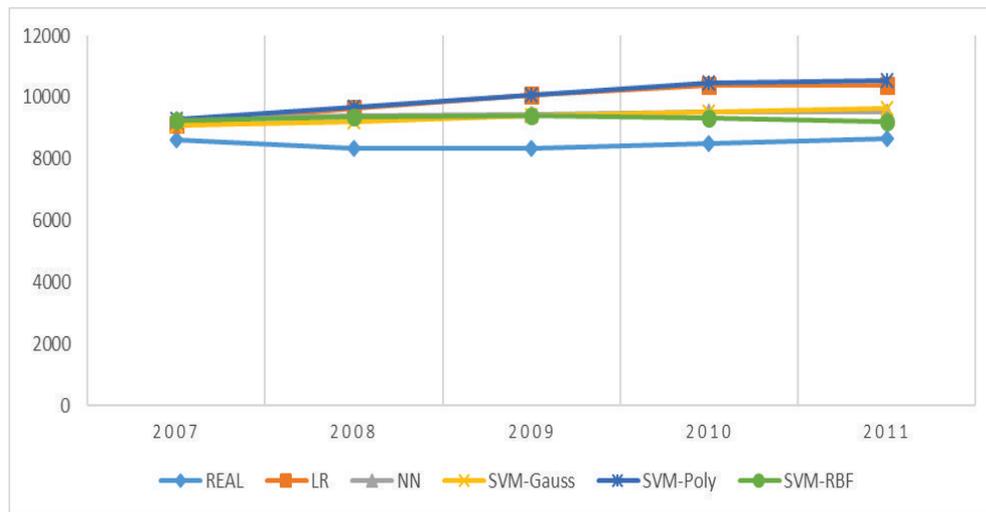


FIGURA 4.23: Predicciones obtenidas para Primaria

Nuevamente podemos basarnos en las métricas para saber qué tan distantes son las variables a los datos reales. Se tiene que según el MAE el menor valor lo tiene RNA con 2.46 puntos seguido de SVM-Pol con 2.97, también RL tiene un buen indicador no tan lejos de los dos anteriores, sin embargo, SVM-Gauss y SVM-RBF si tienen una diferencia más notable. RRSE por su parte tiene mejores números en RNA con 1.13 y sigue SVM-Pol con 1.39 valores muy cercanos, de allí les sigue RL y al final los otros dos modelos con 53 puntos aproximadamente. RAE favorece también a RNA con 0.85 puntos, diferenciándolo de los otros dos indicadores con valores cercanos como son SVM-Pol y RL que tienen una puntuación por encima de 1; Los dos modelos restantes están entre 40 y 50 puntos. MAPE tiene un valor aproximado de 2.74% en RNA que es la mejor predicción según esta métrica, los valores de SVM-Gauss y RL son buenas también con un porcentaje poco mayor del 3%; No así con SVM-RBF y SVM-Gauss que oscilan alrededor del 100% de error. Para el RMSE tenemos un indicador que se inclina en favor de SVM-Pol con 33.24 puntos, RNA y RL no están lejos de esa cifra, pero son poco mayores, SVM-RBF es por otro lado mucho mayor junto con SVM Gauss. Para terminar, tenemos a MSE inclinándose también por SVM-Pol con el indicador mínimo de 10.5 seguido de RNA con más de 11 unidades y RL con 14; Podemos concluir que el

mejor modelo que más se ajusta a los datos reales es el modelo de Redes Neuronales, con cuatro indicadores favoreciéndolo contra dos indicadores que favorecen SVM-Pol aunque en realidad las cifras en las que varían no son tan altas.

CUADRO 4.4: Métricas de comparación para el municipio de Tlanepantla (Primaria)

Métrica	LR	NN	SVM-Gauss	SVM-Poly	SVM-RBF
Mean absolute error	3.3452	2.4636	101.0894	2.9787	77.4225
Root relative squared error	1.5797	1.1326	53.0144	1.3965	53.11
Relative absolute error	1.5428	0.8542	49.8285	1.3959	41.3814
Mean absolute percentage error	0.0361	0.0274	1.0988	0.0325	0.8331
Root mean squared error	3.7681	3.4102	119.208	3.2473	117.2967
Mean squared error	14.1989	11.6294	14210.5578	10.5452	13758.5194

Secundaria tiene en sus datos Reales más parecido a SVM-Gauss y SVM-RBF que han resultado los menos precisos en el caso de predicción para Pre Primaria y Primaria, RNA Muestra un comportamiento cercano a los valores reales, salvo en uno de los periodos. Esta ocasión a simple vista pareciere que SVM-Pol y RL que han tenido buenos indicadores en los niveles educativos anteriores, se aprecia que se encuentran más retirados de la línea Real.

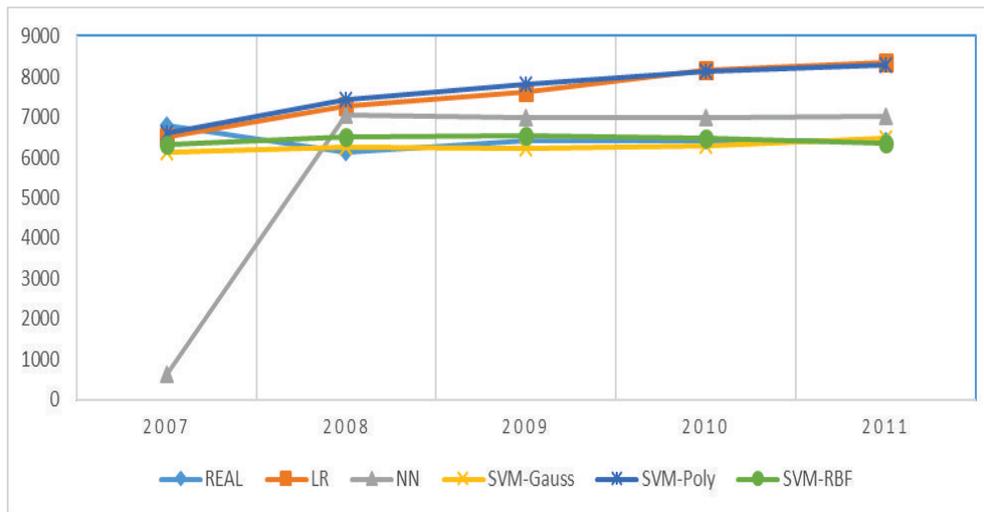


FIGURA 4.24: Predicciones obtenidas para Secundaria

De acuerdo a las métricas podemos leer que en los indicadores MAE el indicador con un nivel de error menor es RNA, que supera claramente el resto de valores en los demás modelos de predicción, con un valor menor a uno de 0.78, mientras el resto

tiene valores más altos de más de 4 puntos y de más de 100 puntos en el caso de SVM-Gauss y SVM-RBF. RRSE también tiene una importante inclinación por RNA con un valor de 0.25 frente a 1.5 aproximadamente de RL y SVM-Pol, las otras dos SVM tienen valores alrededor de 60. RAE también favorece mucho a RNA el único con valor menor a uno, los dos más cercanos siguen siendo RL y SVM-Pol con valor de más de uno, el resto incrementa en mucho más con más de 50 puntos de error. MAPE marca a RNA con 1% y le siguen con 7% y 8% en SVM-Pol y RL respectivamente. Mientras que las dos máquinas SVM oscilan entre 150% y 240% de error. RMSE que hemos visto en los dos casos anteriores favorece a SVM-Pol sobre RNA en este caso no sucede así, RNA es indiscutiblemente el más pequeño indicador de los modelos de predicción, con un valor de 0.8 mientras que los demás son en SVM-Pol de más de 4 y en RL de más de 5 unidades, el resto de modelos presenta un error cercano a las 180 unidades. MSE por otro lado dice que el valor de RNA es menor a uno también, el resto son mayores, de más de 20 unidades en los dos más cercanos para RL y SVM-Pol, y más de 30 mil para los dos modelos restantes.

CUADRO 4.5: Métricas de comparación para el municipio de Tlanepantla (Secundaria)

Métrica	LR	NN	SVM-Gauss	SVM-Poly	SVM-RBF
Mean absolute error	4.9295	0.7894	143.2092	4.3596	92.1846
Root relative squared error	1.6862	0.2547	62.8959	1.4578	59.1354
Relative absolute error	1.8359	0.2849	58.23	1.5733	39.6637
Mean absolute percentage error	0.0829	0.0137	2.4153	0.0754	1.5198
Root mean squared error	5.3451	0.8271	186.1878	4.7038	173.5709
Mean squared error	28.5705	0.6842	34665.9043	22.126	30126.8508

Los valores en Bachillerato nos dicen que el comportamiento de la línea Real es oscilante, el resto de las líneas algunas muestran grandes distancias a la línea de valores reales, mientras el resto parece estar más cerca, aunque la mayoría de puntos predichos por los modelos son mayores a los datos Reales.

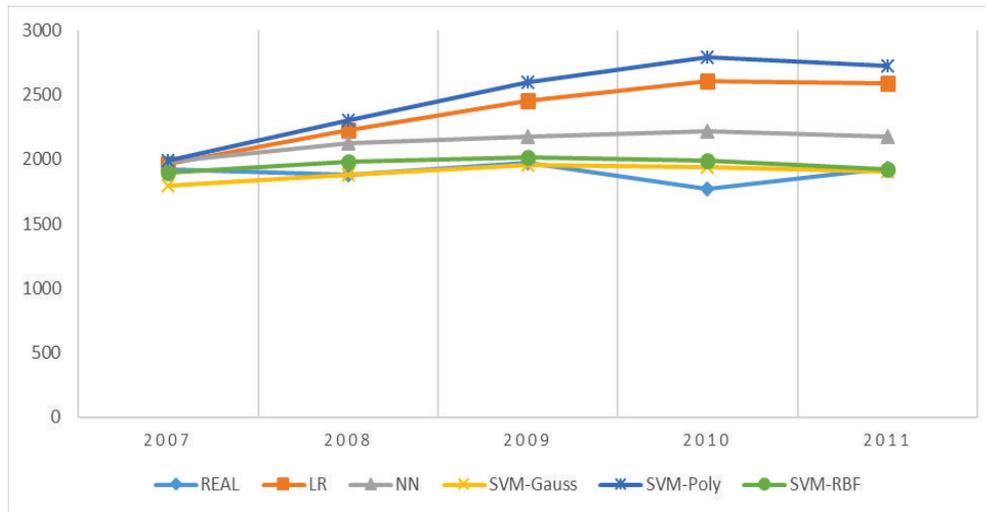


FIGURA 4.25: Predicciones obtenidas para Bachillerato

Según los indicadores de las diferentes métricas también MAE beneficia a RNA con puntaje inferior a uno, los más cercanos son nuevamente SVM-Pol y RL que tienen un puntaje de más de 2, mostrando una cercanía a los valores de RNA, también tenemos datos muy distintos en SVM-Gauss y en SVM-RBF con 70 y 95 puntos respectivamente, valores más alejados que los demás por una cantidad importante. También RRSE nos indica que RNA es un modelo certero con un valor inferior a 1, de hecho, para todas las métricas aparece RNA con valores inferiores a 1, en el resto de las predicciones tenemos otros errores mayores a uno, salvo en dos casos que se verán más adelante. Los valores de RRSE de RL y SVM-Pol tienen valores cercanos de alrededor de 2 puntos con un puntaje mejor ajustado en SVM-Pol; SVM-Gauss y SVM-RBF se disparan nuevamente a valores entre 70 y 85 puntos. RAE presenta un error de 0.37 en las RNA, de más de 2 en SVM-Pol y RL los más ajustados después del RNA, sin embargo, el resto de modelos sigue estando lejos de los demás con puntajes alrededor de 95. MAPE que se mide en porcentajes nos arroja un 1.7% de error en RNA seguido de 13% y 14% en SVM-Pol y RL respectivamente, estos vienen siendo los casos en que el valor es menor que 1. El resto de valores de MAPE son de 400% y 500%. RMSE tiene a RNA como el único por debajo de uno con 0.31 puntos, mientras que sus competidores más cercanos alcanzan los 2 puntos, el resto de modelos competidores hasta 81 y 127 en el caso de SVM-Gauss y SVM-RBF. Por último está el MSE que asigna un valor de 0.098 a RNA y de 5.14 y 6.12 a SVM-Pol y RL respectivamente, SVM-Gauss y SVM-RBF saltan a 6000 y 16000 puntos en ese mismo orden.

Los datos de la serie Real tienen una caída y después un pico alto para volver a

CUADRO 4.6: Métricas de comparación para el municipio de Tlanepantla (Bachillerato)

Métrica	LR	NN	SVM-Gauss	SVM-Poly	SVM-RBF
Mean absolute error	2.371	0.2984	70.0207	2.1854	95.6004
Root relative squared error	2.0119	0.2549	70.3166	1.9104	84.9596
Relative absolute error	2.9886	0.3733	95.3504	2.8552	94.0611
Mean absolute percentage error	0.1405	0.0178	4.0687	0.1304	5.7189
Root mean squared error	2.4743	0.3133	81.2295	2.2689	127.9261
Mean squared error	6.1222	0.0981	6598.2321	5.148	16365.0893

descender, parece que RL y SVM-RBF son aparentemente las series más separadas de los datos Reales, SVM-Pol se nota más cerca o ajustado, SVM-Gauss se parece a la predicción de SVM-RBF, por ultimo RNA muestra un comportamiento más centralizado respecto a los altibajos de los datos Reales.

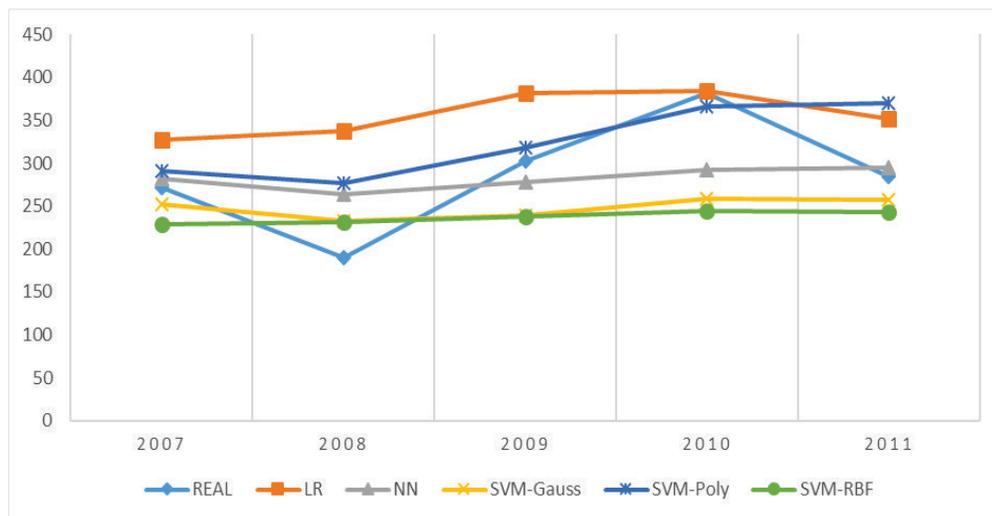


FIGURA 4.26: Predicciones obtenidas para Profesional

Ahora de acuerdo a las métricas tenemos los siguientes datos, hay buenos ajustes de tres series, es importante tomar el resultado de los indicadores sobre evaluación de error porque para eso son útiles en medición cualitativa. El primer indicador MAE nos dice que RL es la mejor predicción de entre los cinco modelos, tiene un puntaje de 0.213, después de este sigue la predicción del SVM-Pol con 0.217, bastante cercano al anterior; enseguida viene el modelo RNA con 0.22 poco mayor, por ultimo tenemos SVM-Gauss con más de 11 puntos y SVM-RBF con casi 29 puntos. Es por eso que decimos que tres series se ajustan mejor y con buenos resultados a los datos Reales. Posteriormente tenemos al RRSE con buenos resultados para RL con 0.59 y tiene un excelente ajuste,

después le siguen RNA y SVM-Pol en ese orden con un valor mayor a 0.6, los resultados más alejados los encontramos con SVM-Gauss en casi 40 puntos y en SVM-RBF con más de 88 puntos. RAE por su parte, también favorece a RL con 0.59, esta vez SVM-Pol es mejor que RNA cada uno con valores mayores a 0.6, los otros dos modelos restantes tienen valores mucho mayores como 36 y 74 puntos respectivamente. MAPE nos muestra los tres mejores modelos que hemos venido viendo están en 11 %, RL con 11.41 %, SVM-Pol con 11.58 % y RNA con 11.77 %, excelentes cifras, SVM-Gauss más de 500 % y SVM-RBG con más de 1300 %. RMSE dice que RL es el de menor cifra con 0.236 seguido inmediatamente de RNA con 0.238 y de SVM-Pol con 0.245, los otros dos modelos SVM-Gauss y SVM-RBF crecen hasta 213 y 1338 respectivamente alejándose considerablemente de los otros tres modelos. Podemos aseverar que el mejor modelo en predicción para Profesional-Técnico es Regresión Lineal. De todo lo anterior resumimos la siguiente información:

CUADRO 4.7: Mejores resultados de las técnicas utilizadas

Nivel académico	Mejores valores en orden		
	1°	2°	3°
Pre-Primaria	RNA/SVM-Pol	RL	SVM-Gauss
Primaria	RNA	SVM-Pol	RL
Secundaria	RNA	SVM-Pol	RL
Bachillerato	RNA	SVM-Pol	RL
Profesional-Tecnico	RL	RNA/SVM-Pol	SVM-Gauss

Por lo tanto, tenemos que el modelo que mejor ajuste tuvo a los datos Reales del Municipio de Tlalnepantla es el modelo de Redes Neuronales artificiales, aunque los otros dos modelos que también se acercaron SVM-Pol y RL, modelos que también generaron en general excelentes aproximaciones a la serie de datos Reales; Según las métricas de evaluación de error el modelo que menos error tiene en sus predicciones para los años 2007 a 2011 en Tlalnepantla es el modelo de Redes Neuronales en la mayoría de los casos. El primer caso de Pre-Primaria tiene un “empate” entre los modelos de RNA y SVM-Pol ya que tienen de seis indicadores tres a favor de RNA y tres de SVM-Pol de allí se sigue RL que presenta una importante cercanía los otros dos, SVM-Gauss entro en la tabla como tercer competidos por causa de que dos modelos empataron en el primer lugar, y no precisamente por que demuestre ser competitivo contra los tres primeros, pues sus valores de error fueron mucho mayores al de los primeros tres modelos. También en el nivel de Profesional-Técnico encontramos que Regresión Lineal lidera la tabla, en todos los indicadores obtiene el mejor puntaje al momento de medir

el error, RNA y SVM-Pol nuevamente vuelven a empatar en segundo lugar con tres indicadores en favor de cada uno de ellos, por ultimo SVM-Gauss vuelve a entrar por razón de posición y no de representar ser competitivo en sus predicciones contra el resto de los tres primeros modelos.

CUADRO 4.8: Métricas de comparación para el municipio de Tlanepantla (Profesional)

Métrica	LR	NN	SVM-Gauss	SVM-Poly	SVM-RBF
Mean absolute error	0.2133	0.2219	11.377	0.217	28.929
Root relative squared error	0.5901	0.6305	39.4092	0.652	88.3284
Relative absolute error	0.5954	0.6607	36.0593	0.653	74.0457
Mean absolute percentage error	0.1141	0.1177	5.2627	0.1158	13.2736
Root mean squared error	0.2361	0.2389	14.6199	0.2454	36.5845
Mean squared error	0.0557	0.0571	213.7418	0.0602	1338.4283



# Capítulo 5

## Conclusiones

En esta investigación se ha desarrollado un trabajo basado en los objetivos inicialmente planteados. Se planteó la posibilidad de aplicar modelos basados en Inteligencia Artificial para poder conocer la situación académica de la población que habita los municipios conurbados del Estado de México; Poder saber en base a variables socio-demográficas, económico-financieras, y otras más, la situación de alumnos egresados de los diferentes niveles educativos, pre-primaria, primaria, secundaria, bachillerato y profesional o técnico, esto da un panorama general de la cantidad de alumnos que terminan sus estudios. Se utilizaron variables o información obtenida de censos y conteos poblacionales, del Banco de Información Económica (BIE) y otras bases de datos en INEGI, así como de otras instituciones gubernamentales. Una vez que se aplicaron procedimientos de limpieza y estructuración de los datos se llevó a cabo una selección de las variables más adecuadas para la investigación realizando esto por medio de Algoritmos Genéticos se obtuvo un espacio más reducido de variables que se pudieron utilizar en los modelados finales. Las técnicas empleadas para modelar la cantidad de egresados de cada nivel fueron Regresión Lineal, Redes Neuronales Artificiales, Maquinas de Vectores Soporte con Kerneles Gaussiano, Polinomial y RBF, utilizando dichas técnicas se generaron predicciones sobre los municipios estudiados con un periodo de cinco años posteriores a los que se tenían con datos reales, que son de 1994 a 2011, por lo tanto, las predicciones fueron de 2012 a 2016. De estas predicciones se analizó el comportamiento de cada una, como se asemejaban o diferenciaban entre ellas. Por último, se realizó un comparativo de los años 2007 a 2011 donde se compararon las predicciones generadas por cada uno de los modelos de Inteligencia Artificial con los datos reales de allí se pudieron obtener la precisión de predicción o ajuste que tiene cada estimación hacia los datos verdaderos. De los análisis realizados y las hipótesis inicialmente planteadas podemos concluir lo siguiente: Es posible la realización de un modelos de Inteligencia Artificial, en este caso Algoritmos Genéticos, para la selección

de mejores características o variables para encontrar las que más y mejor información puedan aportar al momento de realizar predicciones con dicha información; Las variables tomadas de entre una cantidad medida en cientos de variables pueden reducirse a un número mucho menor y que aportan con una excelente precisión información útil sin necesidad de utilizar los cientos de variables iniciales. También en base a esta investigación se asume que se pueden aplicar modelos basados en Inteligencia Artificial para realizar buenas predicciones, basándose en las variables escogida por otro modelo de IA (Algoritmos Genéticos en este caso), de estudiantes de todos los niveles educativos desde preescolar hasta nivel superior con gran nivel de predicción y buen ajuste a las series de datos reales sobre el mismo asunto. Por último, se defiende con lo realizado, que de acuerdo a las métricas de evaluación de errores aplicadas a las predicciones realizadas con los modelos de IA, se puede determinar qué modelo es el mejor para la predicción de cantidad de Egresados de las escuelas en los diferentes niveles educativos en los municipios conurbados del Estado de México. Del mismo modelo se concluye basado en el último punto que según la investigación Los cinco modelos de IA aplicados a los datos, son muy buenos al momento de realizar predicciones y ajustes a los datos reales, pero hay algunos que lo hacen de mejor manera en el siguiente orden:

1. Redes Neuronales Artificiales
2. Maquinas de Vectores Soporte con kernel Polinomial
3. Regresión Lineal
4. Maquinas de Vectores Soporte con kernel Gaussiano
5. Maquinas de Vectores Soporte con kernel Radial Basis Function

De lo anterior se denota que las RNA son el modelo que mejor y más veces se ajustó a los datos reales, así mismo SVM-Pol tuvo un impresionante desempeño no despegándose mucho de los resultados de RNA, por último RL tiene una precisión alta, podríamos decir en ocasiones tan alta como RNA y en ocasiones aun mayor como en el caso de las predicciones del nivel Profesional-Técnico.

Por lo tanto, RNA demuestra ser junto con SVM-Pol una excelente opción al momento de predecir, sin embargo, considerando que RL es un modelo mucho menos complejo matemáticamente que los dos anteriores y su precisión es muy parecida a la de los modelos anteriores, aunque quedo en tercer lugar, sigue siendo una excelente opción y más sencilla al momento de realizar predicciones.

Notas de conclusiones:

2005 aumenta en preescolar porque es el año en que se establece el preescolar como obligatorio RBF es muy lineal y en ocasiones parecido a Gauss RL y SVM Pol son muy parecidos en casi todas las gráficas y tienen tendencias similares RNA muestra comportamientos diferentes al resto de modelos



# Bibliografía

- [(Rodriguez et al, 2014)] Rodriguez Albor G., Ariza Dau M., Ramos Ruiz J.L., Calidad institucional y rendimiento académico, El caso de las universidades del Caribe colombiano, *Perfiles Educativos* vol. XXXVI, núm. 143, 2014, IISUE-UNAM.
- [(Reyes et al. 2014)] Reyes Carreto R., Godinez Jaimez F., Ariza Hernandez F. J., Sanchez Rosas F., Torreblanca Ignacio O. F., Un modelo empírico para explicar el desempeño académico de estudiantes de bachillerato, *Perfiles Educativos* vol. XXXVI, núm. 146, 2014, IISUE-UNAM.
- [(Gomez y Mendoza 2010)] Gómez Gil Ma. del P. y Mendoza Velázquez A., Redes Neuronales Artificiales para Calificar la Capacidad de Crédito de Entidades Mexicanas de Gobierno *Komputer Sapiens*, 2010, Año II, Vol. II., 26 / 32.
- [(Flores 2005)] Flores, J. W. S. Reconocimiento de patrones de morosidad para un producto crediticio usando la técnica de árbol de clasificación CART. Universidad Nacional Mayor de San Marcos. Ingeniería Industrial., 2005.
- [(Pilloneto 2014)] Pillonetto G.F.D., Kernel methods in system identification, machine learning and function estimation: A survey *Pattern Recognition*, Febrero 2014
- [(Petasiuk 2014)] Petasyuk, G. System and criterial method of the identification and quantitative estimation of the geometrical shape of the abrasive powder grain projection *Powder Technology*, Abril 2014
- [(Lakshmanan 2014)] Lakshmanan M.N, Anuj J. Kapadia, Pooyan Sahbaee, Scott D. Wolter, Brian P. Harrawood, David Brady, Ehsan Samei An X-ray scatter system for material idenfication in cluttered objects: A Monte Carlo simulation study *Nuclear Instruments and Methods in Physics Research B*, Mayo 2014

- [(Chao 2006)] Li Chao, Yang Yi-Xian, N.X.x. Biometriebased personal identitywt-hentication system and security analysis THE JOURNAL OF CHI-NA UNIVERSITIES OF POSTS AND TELECOMMUNICATIONS, Diciembre 2006
- [(Asraf 2012)] H. Muhammad Asraf, M. T. Nooritawati, M. S. R. A Comparative Study in Kernel- Based Support Vector Machine of Oil Palm Leaves Nutrient Disease Procedia Engineering, 2012
- [(Xu 2011)] Guili Xu, Fengling Zhang, S. G. S. Y. Y. H. M. Use of leaf color images to identify nitrogen and potassium deficient tomatoes Pattern Recognition Letters, Mayo 2011
- [(Revollo 2012)] Revollo, G. N. Revollo, C. D. G. M. P. Clasificación y Extracción de Características en Imágenes de Playa ECIImag, 2012
- [(Cadena 2011)] Cadena, C., Jarqin, M. y Miño, M., "Estado de México Historia Breve", Fondo de Cultura Económica, 2011.
- [(Capeci 1991)] Capeci John, "Credit Risk, Credit Ratings, and Municipal Bond Yields: a Panel Study", National Tax Journal, Vol. 44, No. 4, December, 1991
- [(Chang 1996)] Chang L. Kun, Hart, I.y Kwon, Y., "Hybrid neural network models for bankruptcy predictions", Decision Support Systems, 1996
- [(Lesberg 2006)] Lensberg T., Eilifsen A. y McKee T. E., "Bankruptcy theory development and classification via genetic programming", European Journal of Operational Research, 2006
- [(Lopez 2003)] López, V., "Deuda Pública de Estados y Municipios: Una alternativa para Financiar Proyector de Infraestructura", Indetec, Primera Edición, 2003
- [(Mejia 2011)] Mejía Reyes, P. y L. Rendón Rojas, "La producción del Estado de México en la era de la economía global", 2011
- [(Ratner 2013)] Ratner R. Ricardo, "Análisis del Credit Scoring ", Revista de Administração de Empresas, Vol. 53 No. 3, 2013

- [(Ravi 2007)] Ravi K. P. y Ravi, V., "Bankruptcy prediction in banks and firms via statistical and intelligent techniques – A review", *European Journal of Operational Research*, 2007
- [(GobMex 2011)] "Regiones del Estado de México", Portal del Gobierno del Estado de México, Estado de México, <http://portal2.edomex.gob.mx/edomex/estado/geografiayestadistica/regiones> (Febrero 2014)
- [(Riguzzi 2011)] Riguzzi, P. y Sobrino, L. J. "Historia General del Estado de México", *El Colegio Mexiquense - Gobierno del Estado de México - Biblioteca Mexiquense del Bicentenario*, Toluca, vol. 6.
- [(Yang 1999)] Yang, Z. R., Platt, M. B. y Platt, H. D., "Probabilistic Neural Networks in Bankruptcy Prediction", *Journal of Business Research*, 1999
- [(Zhang 1999)] Zhang G., Y. Hu, M., Eddy, P. B. y C. Indro, D., "Artificial neural networks in bankruptcy prediction: General framework and cross-validation analysis", *European Journal of Operational Research*, 1999
- [(Angulo 2012)] Fischer Angulo, E., (2012), *Modelo para la automatización del proceso de determinación de riesgo de deserción en estudiantes universitarios*, (Tesis de Magister), Universidad de Chile, Chile
- [(Reyes et al 1999)] Norma G. Reyes, Miguel A. Gil, Myriam Soria, Julio C. Mosqueda, Ricardo M. Sánchez e Ismael Pérez., *Detección de la probabilidad de deserción en alumnos de nuevo ingreso en base a factores de riesgo utilizando redes neuronales artificiales*, *Tecnologías y aprendizaje avances en Iberoamérica*, Vol. 2, 52-59
- [(Altujjar 2016)] Yasmeen Altujjar, Wejdan Altamimi, Isra Al-Turaiki y Muna Al-Razgan, (2016), *Predicting Critical Courses Affecting Students Performance: A Case Study*, *Procedia Computer Science*, 82, 65 – 71.
- [(Camacho et al 2011)] Camacho, J., Cortés, J., Micle, R., Sánchez, A., (2011), *Predicting the academic underachievement in a high school in Spain over the next few years: A dynamic modeling approach*, *Mathematical and Computer Modelling*, 57, 1703–1708

- [(Hamsa et al. 2016)] Hashmia Hamsa, Simi Indiradevi, Jubilant Kizhakkethottam, (2016), Student academic performance prediction model using decision tree and fuzzy genetic algorithm, *Procedia Technology* 25, 326 – 332
- [(Pandey y Taruna 2016)] Pandey, M. y Taruna, S., (2016), Towards the integration of multipleclassifier pertaining to the student’s performance prediction, *Perspectives in Science* 8, 364-366.
- [(Scutter 1990)] Scutter S., (1990), Why do students enrol in the physiotherapy course? A study of students at SALT, *Australian Pyphysiotherapy* 36, 23-28.
- [(Leung et al. 2001)] Leung ,Y., Ma, J., y Zhang, W., (2001), A New Method for Mining Regression Classes in Large Data Sets, *IEEE Transactions On Pattern Analysis And Machine Intelligence*, Vol. 23, No. 1, 5-21.
- [(Shahiria et al 2015)] Mohamed Shahiria, A., Husaina, W. y Abdul Rashida, N., (2015), The Third Information Systems International Conference, *Procedia Computer Science* 72, 414 – 422.
- [(Zhang et al 2015)] Zhang, X., Wang, L., Xiang, S. y Liu, C., (2015), Retargeted Least Squares Regression Algorithm, *IEEE Transactions On Neural Networks And Learning Systems*, Vol. 26, No. 9, 2206-2213.
- [(Torra et al 2016)] Monsalve-Torra, A., Ruiz-Fernandez, D., Marin-Alonso, O., Soriano-Payá, A., Camacho-Mackenzie, J. y Carreño-Jaimes, M., (2016), Using machine learning methods for predicting inhospital mortality in patients undergoing open repair of abdominal aortic aneurysm. *Journal of Biomedical Informatics* 62, 195–201.
- [(Su et al 2012)] Su, Y., Gao, X., Li, X. y Tao, D., (2012), Multivariate Multilinear Regression, *IEEE Transactions On Systems, Man, And Cybernetics—Part B: Cybernetics*, Vol. 42, No. 6,1560-1573.
- [(Dong 2015)] Dong, G., (2015), Pattern-Aided Regression Modeling and Prediction Model Analysis, *IEEE Transactions On Knowledge And Data Engineering*, Vol. 27, No. 9, 2452-2465.

- [(Tay et al. 2015)] Tay, D., Loo Poh, C., Van Reeth, E. y Kitney, R., (2015), The Effect of Sample Age and Prediction Resolution on Myocardial Infarction Risk Prediction, *IEEE Journal Of Biomedical And Health Informatics*, Vol. 19, No. 3, 1178-1185.
- [(Haiyin y Zhiqi 2010)] Haiyin, L. y Zhiqi, L., (2010), Prediction Model of China's Population, 705-708
- [(Liu et al. 2015)] Liu, J., Zhang, Y., Xiao, Z., Qiao, T. y Tan, H., (2015), Fast, Finite, Accurate and Optimal WASD Neuronet versus Slow, Infinite, Inaccurate and Rough BP Neuronet Illustrated via Russia Population Prediction, *Sixth International Conference on Intelligent Control and Information Processing*, Wuhan, China, 26-28.
- [(Zhang et al. 2014)] Zhang, Y., Liu, J., Guo, D., Ding, S. y Tan, H., (2014), Power-Activated WASD Neuronet Based Russian Population Estimation, Correction and Prediction, *IEEE International Conference on Computational Science and Engineering*, 1232-1236.
- [(Zhang et al 2015)] Zhang, Y., Liu, J., Yan, X., Qiu, B. y Qiao, T., (2015), WASD Neuronet Prediction for China's Population, *Proceeding of the 2015 IEEE International Conference on Information and Automation*, Lijiang, China, 797-802.
- [(Ying et al 2014)] Mei Ying, J., Li Ran, X., Zhi Ding, H. y Le De, N., (2014), An analysis of the Change in Population Size of Chinese Wa Nationality and Predictive Based on Census Data and Spectrum Population Prediction Software, *Sixth International Conference on Measuring Technology and Mechatronics Automation*, 75-77.
- [(Guo et al 2015)] Guo, D., Zhang, Y., He, L., Zhai, K. y Tan, H., (2015), Chebyshev-Polynomial Neuronet, WASD Algorithm and World Population Prediction from Past 10000-year Rough Data, *27th Chinese Control and Decision Conference (CCDC)*, 1702-1707.
- [(Zhou et al 2014)] Zhou, A., Jin, Y. y Zhang, Q., (2014), A Population Prediction Strategy for Evolutionary Dynamic Multiobjective Optimization, *IEEE TRANSACTIONS ON CYBERNETICS*, VOL. 44, NO. 1, 40-53.

- [(Belciug y Gorunescu 2014)] Belciug, S. y Gorunescu, F., (2014 ), Error-correction learning for artificial neural networks using the Bayesian paradigm. Application to automated medical diagnosis, *Journal of Biomedical Informatics* 52, 329–337.
- [(Ebrahimabadi et al. 2015)] Ebrahimabadi, A., Azimipour, M. y Bahreini, A., (2015), Prediction of roadheaders' performance using artificial neural network approaches (MLP and KOSFM), *Journal of Rock Mechanics and Geotechnical Engineering* 7, 573-583.
- [(Gutierrez 2011)] Antonio Gutiérrez, P., Hervás-Martínez, C. y Martínez-Estudillo, F., (2011), Logistic Regression by Means of Evolutionary Radial Basis Function Neural Networks, *IEEE Transactions On Neural Networks*, Vol. 22, No. 2, 246-162.
- [(Vapnik 1998)] Vladimir N. Vapnik, *Statistical Learning Theory*, Wiley-Interscience, 768, 1998.