

Software De Proyecciones Y Tendencias De Riesgo En Educación Media En La Entidad Territorial San José De Cúcuta.

Software Of Projections And Trends Of Risk In Middle Education In The Territorial Entity San José De Cúcuta.

Miguel Rodríguez^{1*} 
Universidad de Pamplona

Jose G. Chacón² 
Fuerza Aérea Colombiana

Joseph N. Sequeda³ 
Universidad Simón Bolívar

© 2021 Universidad de Córdoba. Este es un artículo de acceso abierto distribuido bajo los términos de la licencia Creative Commons Attribution License, que permite el uso ilimitado, distribución y reproducción en cualquier medio, siempre que el autor original y la fuente se acreditan.

¹ Ing. Sistemas, Egresado UNIPAMPLONA, Villa del Rosario, Colombia

² *Msc Ciencias de la computación, Profesor TCO, Villa del Rosario, Colombia, Telf. 315396425, Correo: jose.chacon@unipamplona.edu.co

³ Ing. Sistemas, Egresado UNIPAMPLONA, Villa del Rosario, Colombia

RESUMEN

La investigación acerca de las proyecciones y tendencias en la educación media es importante a nivel institucional y nacional porque se pueden reformular e implementar nuevas metodologías de aprendizaje que sean más eficientes con las nuevas modalidades de estudio. En la presente investigación, se propone desarrollar un software de proyecciones y tendencias de riesgo en educación media utilizando la data del Ministerio de Educación Nacional como apoyo a decisiones en la entidad territorial de san José de Cúcuta. Los datos utilizados en la investigación son los datos registrados en los últimos nueve años de la educación media. Se utilizó la metodología Scrumban en sus diferentes fases para su construcción por las características propias de la investigación. Finalmente, es importante señalar que los resultados esperados de este software buscan dar apoyo en las decisiones que contribuirán a disminuir el índice de repetición, reprobación y deserción en la educación media en el municipio de san José de Cúcuta.

PALABRAS CLAVE: Proyecciones, Tendencias, Metodología Scrumban, riesgo en educación, pronóstico.

ABSTRACT

Research on projections and trends in secondary education is important at the institutional and national level because new learning methodologies can be reformulated and implemented that are more efficient with the new study modalities. In the present research, we propose to develop a software of projections and tendencies of risk in secondary education using the data of the Ministry of National Education as support to decisions in the territorial entity of San José de Cúcuta. The data used in the research are the data registered in the last nine years of high school education. The Scrumban methodology was used in its different phases for its construction due to the characteristics of the research. Finally, it is important to point out that the expected results of this software seek to support decisions that will contribute to decrease the repetition, failure and desertion rates in high school education in the municipality of San José de Cúcuta.

KEYWORDS: Projections, Trends, Scrumban Methodology, risk in education, prognosis

INTRODUCTION

A nivel institucional y nacional es importante resaltar investigaciones en el área de regresiones y proyecciones en la educación, ya que gracias a los resultados obtenidos en estas, se pueden reformular los modelos educativos existentes e implementar nuevas metodologías de aprendizaje que sean más eficientes con las nuevas modalidades de estudio y de igual forma nos ayudan a comprender los motivos

por el cual la educación se debe ir acoplando a las nuevas retos y necesidades de los estudiantes, con la finalidad de impulsarlos al crecimiento personal y social para sus siguientes niveles educativos. Es importante resaltar el trabajo realizado en (Mitre y Hernández 2014) donde se pronostican las tecnologías más prometedoras que probablemente impactarán en la educación en tres horizontes; el corto plazo (el año del informe), el medio plazo (los próximos 2 años) y el largo plazo (los próximos 4 años). Este documento analiza la evolución de las

tendencias tecnológicas de 2004 a 2014, que corresponden a las predicciones a largo plazo del Informe Horizon más reciente. El estudio analiza a través del análisis bibliométrico qué tecnologías tuvieron éxito y se convirtieron en una parte regular de los sistemas educativos, cuáles no tuvieron el impacto previsto y por qué, y la forma de los flujos de tecnología en los últimos años.

Igualmente importante es el trabajo de (Rodrigo 2020). que define las proyecciones como métodos de predicción en los que partiendo de determinadas series de datos, se formula una proyección a futuro con el objetivo de evaluar la ocurrencia probable de cualquier acontecimiento o el desarrollo de una tendencia, esto con el objetivo de predecir el desarrollo futuro de un sistema para ayudar a la toma de decisiones (de planificación) sobre medidas de apoyo, contramedidas u otras acciones que influyan sobre la tendencia del objeto planificado, De igual forma (Estebana 2011) define las herramientas para tendencias como una herramienta web que automatiza los cálculos para la obtención de distintos indicadores poblacionales importantes para el control de enfermedades o eventos de la salud. Se estructura en cuatro módulos; a) una descriptiva que incluye el cálculo del porcentaje, el número de casos, la tasa cruda, la tasa ajustada, la tasa truncada y la tasa acumulada; b) la estimación del porcentaje de cambio anual de las tasas; c) el cálculo de casos esperados, y d) la razón de incidencia o mortalidad estandarizada. La aplicación solicita unos parámetros de entrada al usuario. Una vez procesados los datos y obtenidos los resultados, éstos se envían por correo electrónico al usuario. Los resultados se obtienen para cada una de las causas de estudio (enfermedades, etnias, zonas geográficas...) y cada uno de los sexos introducidos en el fichero base.

Cabe mencionar el trabajo de (Oei 2020) que escribe que el riesgo educativo es la consecuencia de factores concurrentes que definen un contexto de inequidad e impiden, a niños y adolescentes acceder a una educación de calidad, es decir, apropiada a sus necesidades y expectativas y que además asegure igualdad de oportunidades de desarrollo personal, inserción social y laboral y el acceso exitoso al tercer nivel educativo, si la persona lo desea. No menos importante es el trabajo de (Yacuzzi 2020) donde escribe que por pronóstico se entiende el conjunto de actividades a través de las cuales, a partir de datos históricos relevados del entorno (series cronológicas, experiencia cualitativa), se obtienen escenarios y proyecciones de los valores futuros de las variables bajo análisis. Para ello se utilizan modelos econométricos y modelos de series cronológicas.

Tomando como referencia los trabajos anteriores, se desarrolló un software de proyecciones y tendencias de riesgo en educación media utilizando la data del Ministerio de Educación Nacional como apoyo a decisiones en la entidad territorial de san José de Cúcuta. Los datos utilizados en la investigación pertenecen a los periodos académicos del 2011 al 2019, los cuales en su mayoría son de tipo decimales y alfanuméricos. Se utilizó una adaptación de la metodología Scrum para el desarrollo del software implementado sus diferentes fases, tales como; la planificación de la iteración y selección de requisitos, ejecución de la iteración, ajuste de estándares del producto, revisión de Sprint y cierre. Esta es una metodología ágil que nos permite tener una flexibilidad en la gestión y planificación, llevar un seguimiento y registro a lo largo de la evolución del proyecto y una priorización adecuada de las tareas.

1. MATERIALES Y MÉTODOS

Se utilizo la metodología de desarrollo Scrumban (Perez 2020). Esta es una metodología de desarrollo ágil híbrida, que presenta características pertenecientes a las metodologías Scrum y Kanban, por lo tanto, cogiendo lo mejor de Scrum y Kanban, nace Scrumban la cual permite a los equipos moverse hacia el desarrollo Agile y a mejorar constantemente sus procesos. Scrumban tiene como características a destacar, la reducción del tiempo dedicado a planificación y estimación, promoviendo el flujo constante de trabajo, propone una cultura de mejora continua haciendo hincapié en la entrega continua sin sobrecargar al equipo de desarrollo, a su vez alienta la filosofía de “pensar en grande, actuar en pequeño, fallar rápido, aprender rápidamente”, el diagrama de la metodología Scrumban se muestra en la figura 1.

Para la ejecución de la primera fase de la metodología Scrumban donde se planifica la lista de tareas y se selecciona los requisitos de las iteraciones, esta planificación se desarrolla en la tabla 1 donde se muestra el nombre de la tarea, el nivel de detalle, la prioridad que va desde 1 siendo la mayor prioridad y 5 la menor, y el periodo de ejecución que esta dado en semanas. Se seleccionaron los datos de la base de datos del Ministerio de Educación y se organizan en 40 % de los datos para entrenar y el 60 % de los datos para validar. Estos se muestran en la tabla 2. Posteriormente se crean los modelos de regresión Lineal, Ridge y Lasso. La regresión lineal simple consiste en generar un modelo de regresión (ecuación de una recta) que permita explicar la relación lineal que existe entre dos variables. A la variable dependiente o respuesta se le identifica como Y y a la variable predictora o independiente como X (Cienciadedatos 2020).

El modelo lineal se define con sklearn `linear_model Linear Regression`: se ajusta a un modelo lineal con coeficientes $w = (w_1, \dots, w_p)$ para minimizar la suma residual de cuadrados entre los objetivos observados en el conjunto de datos y los objetivos predichos por la aproximación lineal.

La regresión Ridge es similar al ajuste por mínimos cuadrados en cuanto que ambos tratan de minimizar el Residual Sum of Squares (RSS). La diferencia reside en que Ridge Regression incorpora un término llamado shrinkage penalty que fuerza a que los coeficientes de los predictores tiendan a cero. El efecto de esta penalización está controlado por el parámetro λ . Cuando $\lambda=0$ la penalización es nula y los resultados son equivalentes a los obtenidos por mínimos cuadrados, cuando $\lambda=\infty$ todos los coeficientes son cero, lo que equivale al modelo sin ningún predictor (modelo nulo). El modelo Ridge se define con sklearn `linear_model Ridge`: Ridge aborda algunos de los problemas de los mínimos cuadrados ordinarios imponiendo una penalización en el tamaño de los coeficientes con una regularización de L2, Rodrigo J (2020).

El método Lasso es una alternativa al ajuste por Ridge regression que permite superar su principal desventaja, la incapacidad de excluir predictores del modelo. El método lasso, al igual que Ridge regression, fuerza a que las estimaciones de los coeficientes de los predictores tiendan a cero. A diferencia de Ridge, Lasso sí es capaz de fijar algunos de ellos exactamente a cero, lo que permite además de reducir la varianza, realizar selección de predictores. Como resultado, el método Lasso tiende a generar modelos más fáciles de interpretar que los obtenidos mediante Ridge regression. A esto se le conoce como sparse modeling El modelo Lasso se define con sklearn `linear_model Lasso`: El Lasso es un modelo lineal que estima los coeficientes de dispersión con una regularización de L1.

Para el entrenamiento de cada uno de los modelos se utilizaron los parámetros de entrada los datos de entrenamiento y la función fit de los modelos, Rodrigo J (2020). Rodrigo J (2020). (AAR Gómez, AV Ibáñez 2017)

Una vez los modelos estén entrenados podemos avanzar con el pronóstico y para ello utilizaremos los datos de validación y la función predict que nos retorna la variable predictiva para los periodos de prueba. Asimismo, podemos definir los criterios estadísticos de validación de los modelos. Con los datos suministrados anteriormente e implementado la librería Matplotlib se generan las gráficas de regresiones de las tendencias de riesgo. Para la construcción del proyecto el cual debe ser en un entorno web por sus propias especificaciones, se implementó el editor de código fuente Visual Code y el lenguaje de programación Python efectuado con el framework Flask. (Desarrolloweb 2020)

Coefficiente de determinación

El coeficiente de determinación es una medida estadística de la bondad del ajuste o fiabilidad del modelo estimado a los datos. Se representa por R^2 e indica cuál es la proporción de la variación total en la variable dependiente (Y), que es explicada por el modelo de regresión estimado, es decir, mide la capacidad explicativa del modelo estimado.

El coeficiente de determinación no solo mide la capacidad explicativa de un modelo, sino que, además, permite elegir entre varios modelos cuál es el más adecuado. Así si los modelos tienen la misma variable dependiente y el mismo número de variables explicativas, será más adecuado el que tenga un coeficiente de determinación mayor.

Los métodos estadísticos utilizados fueron : a) Suma de cuadrados de los

residuos (SCE) La suma de cuadrados de los residuos (SCE) es la parte de la variabilidad de la variable dependiente que no conseguimos explicar con el modelo. Es la parte que nuestro conjunto de variables independiente no consigue explicar de la variable dependiente, en palabras más simples, la suma de cuadrados de los residuos (SCE) representa con una cifra lo que un modelo no es capaz de explicar. Economipedia (2020). b) El coeficiente de determinación esta es una medida estadística de la bondad del ajuste o fiabilidad del modelo estimado a los datos. Se representa por R^2 e indica cuál es la proporción de la variación total en la variable dependiente (Y), que es explicada por el modelo de regresión estimado, es decir, mide la capacidad explicativa del modelo estimado. El coeficiente de determinación no solo mide la capacidad explicativa de un modelo, sino que, además, permite elegir entre varios modelos cuál es el más adecuado. Así si los modelos tienen la misma variable dependiente y el mismo número de variables explicativas, será más adecuado el que tenga un coeficiente de determinación mayor, (Wolters 2020).

2. RESULTADOS

A continuación, pasamos con las interfaces de proyección de reprobación, repitencia que corresponden a las clases pro_reprobacion.html, pro_desercion.html y pro_repitencia.html que se muestran en la figura 2 y figura 3 respectivamente.

A continuación, se desarrolló la evaluación de las regresiones de los modelos de reprobación, deserción y repitencia. Para la validación se realizaron quince iteraciones para cada uno de los modelos, en donde se compararon los resultados para seleccionar el modelo más adecuado. Luego seleccionamos la mejor iteración para cada uno de los modelos según el valor de la

varianza explicada, el cual entre más cercano a uno tendrá una mejor predicción y explicación del modelo, las iteraciones seleccionadas se evidencian en la tabla 3.

Según el modelo seleccionado para la regresión de reprobación, se evidencia que el mayor valor para la varianza explicada es de 0,9544 el cual pertenece a la regresión Lasso, se realizó su respectiva proyección en donde se evidencia un aumento en la tasa de reprobación para los periodos del 2020 al 2024.

3. DISCUSIÓN

Se obtuvieron resultados de la regresión y proyección de reprobación, resultados de la la regresión y proyección de deserción y resultados de la la regresión y proyección de repitencia se evidencia que el mayor valor para la varianza explicada es de 0,9544 el cual pertenece a la regresión Lasso siendo esta la más precisa en cuanto a regresiones y proyecciones donde se evidencia un aumento en la tasa de reprobación para los periodos del 2020 al 2024. Hasta el momento no se registran resultados de otras investigaciones en el mismo campo.

CONCLUSIONES

En cuanto al desarrollo del proyecto y los resultados encontrados se puede concluir que las proyecciones en las tasas de riego en la educación media en el municipio de San José de Cúcuta para los periodos del 2020 al 2024 tienden a un aumento promedio 0.55 en la tasa de reprobación, un descenso promedio 0.095 en la tasa de deserción y un valor constante 1.41 en la tasa de repitencia. De igual forma emplear un entorno de desarrollo que se adapte a las necesidades del investigador y su configuración con

librerías hace que los tiempos y experiencias de desarrollo sean más ágiles y agradables logrando mayor productividad. Como ejemplo está Visual Code,

Finalmente implementar una metodología de desarrollo Scrumban, es de vital importancia en la ejecución de proyectos a corto plazo y que requieran una correcta interpretación de requerimientos o historias de usuario para la mejor implementación de interfaz gráfica de usuario en la web (GUIs) a la medida del usuario final.

REFERENCIAS

- [1]. AAR Gómez, AV Ibáñez. (2017). Aplicación de la Minería de Datos en la educación en línea 2017 ISSN: 1692-7257 - Vol 1 – N 29 - 2017.
- [2]. Cerón y J. Quintero. “Línea de tiempo de la evolución de los MOOC”, en Ingeniería E Innovación, vol. 6, núm. 1, pp. 40-46, 2019 [en línea]. Disponible en: <https://doi.org/10.21897/23460466.1546>
- [3]. Estebana L (2011), «SART (Statistical Analysis of Rates and Trends): herramienta vía web para el cálculo estadístico de indicadores poblacionales,» Gaceta Sanitaria, vol. 5, nº 25, pp. 427-431
- [4]. Nañez, J, Solano J, Bernal E, (2018). Actitudes Y Percepciones De Los Estudiantes, Docentes Y Directivos Sobre Enseñanza Y Aprendizaje Flexibles, E Incorporación De Tic.

- Revista Ingeniería e Innovación 6(1). Obtenido de:
Universidad de Tolima. Tolima-Colombia
- [5]. Mitre-Hernández, (2014) «Estimación y control de costos en métodos ágiles para desarrollo de software: un caso de estudio» *Estimation and Control in Agile Methods for Software Development: a Case Study*, ScienceDirect, vol. 15, n° 3, pp. 403-418,
- [6]. Perez A (2019) . la metododolgia Scrumban. Business School. Tomado de <https://www.obsbusiness.school/blog/la-metodologia-scrumban-cuando-y-por-que-utilizarla> Último acceso: 10 03 2019].
- [7]. Rodrigo C. (2020) , «es.slideshare.net,» 15 04 2011. [En línea]. Available: <https://es.slideshare.net/Gustenrq/proyeccion-es-estadsticas-final>. [Último acceso: 10 03 2020].
- [8]. studentplace, «studentplace98.blogspot.com,» 09 09 2018. [En línea]. Available: <https://studentplace98.blogspot.com/2018/09/metodologia-de-desarrollo-de-software.html>. [Último acceso: 12 03 2020].
- [9]. Salas, D. (2016). *Revolución 4.0. Ingeniería e Innovación*, 4(2), 5-6. <https://revistas.unicordoba.edu.co/index.php/r ii/issue/view/131>
- [10]. Oei E (2020), «La-cara-secreta-del-riesgo,» 09 09 2014. [En línea]. Available: <https://www.oei.es/historico/divulgacioncientifica/?La-cara-secreta-del-riesgo>. [Último acceso: 9 03 2020].
- [11]. Vanegas, D., Ramón, A. A. y Lizarazo, A. K. (2017). *Comunidad y cultura ambiental. dinámicas de potenciación para un desarrollo sostenible y corresponsable*. Revista Ambiental Agua, Aire y Suelo. ISSN 1900-9178. Volumen (8), Numero (1).
- [12]. Yacuzzi E (Universidad del CEMA), Guillermo Paggi (Whirlpool Argentina) , «ucema.edu.ar,» 18 02 2002. [En línea]. Available: <https://ucema.edu.ar/publicaciones/download/documentos/209.pdf>. [Último acceso: 11 03 2020].
- [13]. M. Solarte, G. A. Ramírez and D. A. Jaramillo, “Hábitos de ingreso y resultados en las evaluaciones en cursos en línea masivos con reconocimiento académico,” *Ing. E Innov.*, vol. 5, 1, 2017.
- [14]. www.cienciadedatos.net, «www.cienciadedatos.net,» 14 06 2020. [En línea]. Available: https://www.cienciadedatos.net/documentos/24_correlacion_y_regresion_lineal. [Último acceso: 29 10 2020].
- [15]. Rodrigo J (2020), «www.cienciadedatos.net,» 15 12 2016. [En línea]. Available: https://www.cienciadedatos.net/documentos/31_seleccion_de_predictores_subset_selection_ridge_lasso_dimension_reduction. [Último

acceso: 29 10 2020]

[Último acceso: 29 10 2020]

[16]. Desarrolloweb, «Que es Python,» 19 11 2003. [En línea]. Available: <https://desarrolloweb.com/articulos/1325.php>. [Último acceso: 12 03 2020].

[18]. wolters kluwer, «www.wolterskluwer.es,» 29 10 2020. [En línea]. Available: https://guiasjuridicas.wolterskluwer.es/Content/Documento.aspx?params=H4sIAAAAAAAAAEAMtMSbF1jTAAASmJm0sTtbLUouLM_DxbIwMDS0NDA1OQQGZapUt-ckhIQaptWmJOCsoAQ9XICzUAAAA=WKE. [Último acceso: 29 10 2020].

[17]. Economipedia (2020).com, «economipedia.com,» 29 10 2020. [En línea]. Available: <https://economipedia.com/definiciones/suma-de-cuadrados-de-los-residuos-sce.html>.

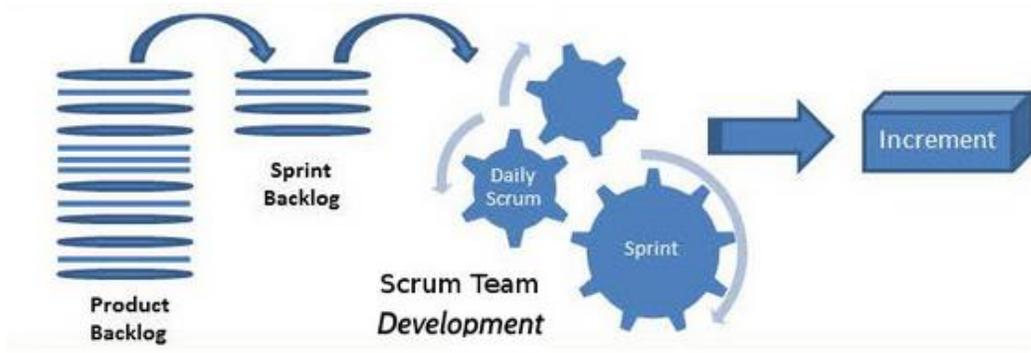


Figura 1: Metodología Scrumban. Fuente: (Pérez 2019).

Tabla 1: Planificación de las iteraciones. Fuente: Autor

| Lista de Tareas | Prioridad | Nivel de Detalle | Periodo de Ejecución (semanas) |
|--|-----------|------------------|--------------------------------|
| Extracción y análisis de la Data MEN. | 1 | Bajo | 0 a 1 |
| Organización y filtración de la Data MEN. | 1 | Bajo | 0 a 1 |
| Identificar y seleccionar los parámetros de riesgo. | 1 | Bajo | 0 a 1 |
| Determinar los procesos estadísticos. | 1 | Alto | 1 a 4 |
| Desarrollo del módulo Index. | 1 | Alto | 8 a 16 |
| Desarrollo del módulo regresión de reprobación. | 2 | Alto | 2 a 4 |
| Desarrollo del módulo regresión de Deserción. | 2 | Alto | 2 a 4 |
| Desarrollo del módulo regresión de repitencia. | 2 | Alto | 2 a 4 |
| Desarrollo del módulo proyecciones de reprobación. | 2 | Alto | 2 a 4 |
| Desarrollo del módulo proyecciones de deserción. | 2 | Alto | 2 a 4 |
| Desarrollo del módulo proyecciones de repitencia. | 2 | Alto | 2 a 4 |
| Desarrollo del módulo Layout. | 3 | Intermedio | 8 a 16 |
| Desarrollo del módulo Importar Data MEN. | 3 | Intermedio | 1 a 2 |
| Desarrollo del módulo graficas de tasas de riesgo. | 4 | Intermedio | 1 a 2 |
| Desarrollo del módulo Información. | 5 | Intermedio | 1 a 2 |
| Desarrollo del módulo regresiones y proyecciones sintéticas. | 5 | Alto | 1 a 3 |
| Desarrollo del módulo de regresiones y proyecciones de aprobación. | 5 | Alto | 1 a 2 |

Tabla 2: Datos relevantes. Fuente: Autor

| Periodos | Tasa de Aprobación | Tasa de Reprobación | Tasa de Deserción | Tasa de Repitencia |
|----------|--------------------|---------------------|-------------------|--------------------|
| 2019 | 91,46 | 6,12 | 2,42 | 1,31 |
| 2018 | 92,44 | 4,54 | 3,01 | 1,06 |
| 2017 | 93,84 | 3,14 | 3,02 | 0,72 |
| 2016 | 93,09 | 3,18 | 3,73 | 1,43 |
| 2015 | 95,16 | 1,11 | 3,73 | 1,26 |
| 2014 | 96,69 | 0,65 | 2,66 | 1,69 |
| 2013 | 95,57 | 0,63 | 3,6 | 1,88 |
| 2012 | 94,5 | 1,15 | 4,35 | 1,3 |
| 2011 | 89,89 | 1,71 | 8,4 | 0,95 |

Tabla 3 Modelos seleccionados. Fuente: Autor

| TABLA DE LOS MODELOS SELECCIONADOS | | | | |
|------------------------------------|-------------|--------------|-----------------------------------|--------------------|
| MODELOS | REGRESIONES | COEFICIENTES | SUMAS DE LOS RESIDUOS AL CUADRADO | VARIANZA EXPLICADA |
| REPROBACION | Lineal | 0,63823 | 1,5457 | 0,8445 |
| | Lasso | 0,55386 | 1,6378 | 0,9544 |
| | Ridge | 0,5784 | 0,4632 | 0,9045 |
| DESERCION | Lineal | 0,6441 | 0,8008 | 0,8253 |
| | Lasso | 0,2489 | 0,3058 | 0,303 |
| | Ridge | 0,4652 | 0,2439 | 0,4441 |
| REPITENCIA | Lineal | 0,036 | 0,1842 | 0,676 |
| | Lasso | 0 | 0,1042 | 0,9476 |
| | Ridge | 1,8 | 0,0843 | 0,777 |

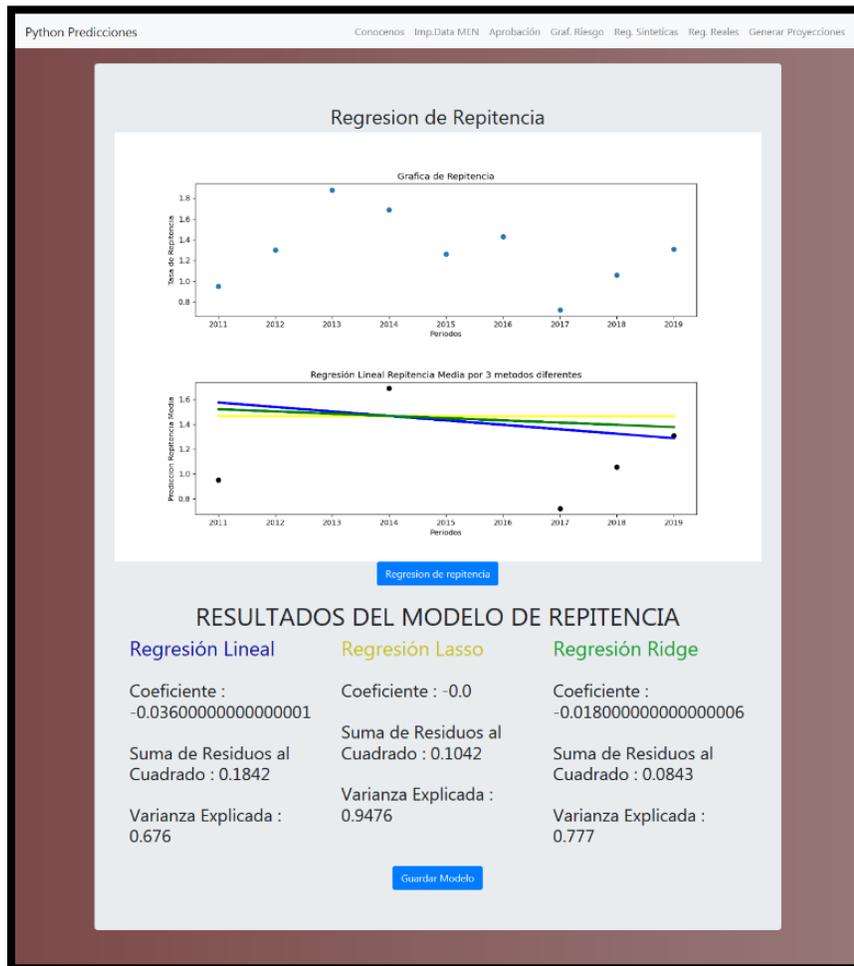


Figura 2 interfaces de proyección de repitencia y resultados. Fuente los autores



Figura 3 Interfaces de proyección de reprobación y resultados. Fuente los autores