

# Clasificación del rendimiento de los estudiantes en la utilización de la plataforma e-educativa en el IES de Charata ingreso 2021

Juan Emiliano Jara

juanemiliano.jara@gmail.com

## Resumen

La pandemia por Covid-19 en el 2020 y 2021 las plataformas virtuales de e-learning han dejado una cantidad de datos que posibilitan la toma de decisión si son correctamente tratados.

En este trabajo se presenta un modelo de clasificación basado en Naïves Bayes y machine learning para detectar el rendimiento en un taller de ingreso dentro un Instituto de Formación Docente en la provincia del Chaco.

Para ello se analiza un conjunto de datos recuperados de los estudiantes ingresantes al ciclo académico 2021 de las siete carreras de formación docente y la tecnicatura superior dictadas en el Instituto de Educación Superior de Charata mediante la plataforma LMS (learning management system o sistema de gestión del aprendizaje) E-educativa.

Palabras claves: Inteligencia artificial, aprendizaje de máquina, formación docente, e-educativa

## 1. Introducción

En el Instituto de Educación Superior de Charata disponemos desde el año 2008 de un nodo de educativa (LMS provisto por Instituto Nacional de Formación Docente), el mismo cuenta con una página web y un campus virtual que se utiliza para el dictado de clases virtuales, hasta el 2019 se llevaron adelante capacitaciones institucionales para que los docentes comiencen a utilizar esta herramienta considerada por algunos de mucho valor para mantener un contacto extra clase con los estudiantes de las distintas carreras del IES.

Con el advenimiento de la pandemia por Covid-19 quedaron al descubierto muchas falencias en sistema educativo de formación docente, el 2020 a la mayoría de docentes golpeo de frente y sin avisar, pese a disponer de las herramientas y algunos de los conocimientos tuvimos que comenzar a trabajar de manera virtual, preparar clases y materiales educativos para presentar a los estudiantes en formato digital y a distancia.

Este desafío se presentó de la noche a la mañana y la mayoría de los docentes estuvo dispuesto a afrontarlo, muchos docentes formaron sus aulas virtuales en e-educativa (INFD), Moodle (ELE plataforma provista por el Ministerio de Educación del Chaco) y Google

ClassRoom; otros docentes decidieron trabajar con grupos cerrados de redes sociales Facebook y una comunicación e intercambio de materiales por WhatsApp.

Durante el inicio académico de este 2021 y por solicitud del Ministerio de Educación, Cultura, Ciencia y Tecnología de la provincia del Chaco los talleres de ingreso a las carreras de profesión docente y tecnicatura profesionales se debían dictar mediante el uso de la plataforma educativa.

Esto brinda una cantidad considerable de información sobre los estudiantes relacionada al uso de las aulas virtuales de los talleres de ingresos y del cursado en los distintos espacios curriculares de los siete profesados y una tecnicatura que cubren un alumnado cercano a los mil quinientos ingresantes, más los cerca de dos mil estudiantes cursantes de años anteriores.

Para lograr la clasificación de rendimiento en la utilización de la plataforma E-ducativa de los estudiantes ingresantes a las carreras de Formación Docente en este 2021 de Instituto de Educación Superior de Charata para una posterior toma de decisión.

## 2. Trabajos relacionados

En [1] se desarrolla una metodología para clasificar y predecir usuarios en ambientes virtuales de educación, estudiando la interacción de los estudiantes con la plataforma y su desempeño en los exámenes. Para esto se utilizaron las herramientas de aprendizaje automático, componentes principales, clusterización, lógica difusa, y el algoritmo del K vecino más cercano. La metodología relaciona los usuarios según las variables de estudio, para así implementar un análisis de clúster que identifica la formación de grupos. Finalmente utiliza un algoritmo de aprendizaje automático para clasificar los usuarios según su nivel de conocimiento.

Mientras que en [2] se describe el desarrollo y evaluación de un Modelo Computacional denominado SocialMining, basado en el algoritmo Naive Bayes, para apoyar el análisis de las opiniones de los estudiantes en el proceso de la evaluación del desempeño docente, llevada a cabo mediante dispositivos móviles. Esta propuesta considera el uso de dispositivos móviles para la recopilación de datos aprovechando su aceptación por parte de los estudiantes en el proceso de educación y aprendizaje.

En [11] proponen una explicación sobre el excelente desempeño de clasificación del algoritmo Naives Bayes Gaussiano. Buscan demostrar si se distribuye la dependencia local de un nodo en cada clase, de manera uniforme o desigual, y cómo las dependencias locales de todos los nodos trabajan juntas, de manera consistente.

### 3. Materiales y datos

#### a. Fuentes de datos

Lo datos del presente trabajo fueron extraídos de la plataforma virtual E-ducativa del ciclo académico 2021 del Instituto de Educación Superior de Charata. E-ducativa es un LMS que permite entre otras cuestiones implementar una plataforma de elearning. Dentro la misma se pueden solicitar reportes que: analizan el acceso por alumno, páginas vistas, recursos utilizados, progresos, tiempos; detallan el progreso de los alumnos en el curso y lectura del material; entregan informes de resultados de exámenes por alumno y aulas; detallan el progreso de los alumnos en el curso y lectura del material, permiten realizar reportes de exámenes e informes de resultados de exámenes por alumno y aulas entre otros tipos de reportes.

Estos datos fueron preparados y depurados para poder utilizarlos en el proyecto, para ello; estandarizaron los datos, normalizaron los datos y realizar una eliminación de columna. Actualmente se dispone un archivo con 756 registros de los cuales se utilizaron solo 381 registros.

#### b. Tratamientos de datos

Nuestra principal fuente de datos son dos tipos de informes que arroja la plataforma e-ducativa (reporte integral de cursado, junto al reporte de lectura y participación), de los informes anteriores se unificaron columnas y se logró el dataset (`iescharata_pru.csv` alojado en GitHub) el mismo está conformado como se detalla a continuación: *Tiempo*; hace referencia al tiempo promedio utilizado en la plataforma. *Avance*; es el porcentaje de ejecución de las distintas actividades en la plataforma. *Realiza*; es el porcentaje de tareas cumplimentadas. *Presenta*; es el porcentaje de actividades presentadas. *Textos*; es el porcentaje de los textos leídos. *Material*; porcentaje de descarga de materiales educativos de la plataforma. *Foros*; porcentaje de participación en los foros. *Aprueba*; el valor cero identifica a los que no aprueban el taller y el valor 1 a los que si aprueban el taller.

	Tiempo	Avance	Realiza	Presenta	Textos	Material	Foros	Aprueba
0	3.72	32	45.27	33.33	22.73	100.00	25.0	1
1	1.00	20	16.00	0.00	18.18	33.33	12.5	0
2	7.45	72	54.55	0.00	68.18	100.00	50.0	1
3	0.15	8	9.47	0.00	4.55	33.33	0.0	0
4	0.00	0	0.00	0.00	0.00	0.00	0.0	0
5	0.00	0	0.00	0.00	0.00	0.00	0.0	0
6	0.25	36	10.23	0.00	40.91	0.00	0.0	0
7	0.10	44	12.50	0.00	50.00	0.00	0.0	0
8	1.13	20	5.68	0.00	22.73	0.00	0.0	0
9	6.22	48	30.21	0.00	50.00	33.33	37.5	1

Fig. 1: Dataset [iescharata\\_pru.csv](#)

Para lograr la tabla de la *Fig. 1* se realizó una reducción de ocho columnas en las distintas fuentes de datos del *reporte integral de cursado* de cada una de las ocho carreras y una reducción de dieciocho columnas de datos específicos dejando los datos con porcentajes totales en los *reportes de lectura y participación* de cada una de las carreras.

Algunos supuestos para el problema que se quiere abordar son; los avances en la realización de las actividades en la plataforma deben ser significativamente mayores de un treinta por ciento, de manera conjunta con la realización de las actividades. Mientras que la participación a los foros debe ser un porcentaje cercano al setenta por ciento al igual que la lectura y descarga de materiales educativos.

En bases a esta información deseamos que el algoritmo aprenda y que se pueda consultar como resultados nueva información que nos brinde la clasificación sobre el desempeño en los talleres de ingreso por parte de los estudiantes de formación docente.

### c. Metodología

#### i. Modelo de aprendizaje de máquina

Se parte de un enfoque supervisado para la toma de decisión utilizando algoritmo de clasificación Naives Bayes Gaussiano, para las métricas de desempeño a utilizar matriz de confusión.

El aprendizaje automático supervisado es la búsqueda de algoritmos que partan de instancias suministradas externamente para producir hipótesis, que luego hacen predicciones sobre instancias futuras, se describen varios algoritmos de clasificación a ser utilizados para el aprendizaje automático supervisado, así como determina los más eficientes algoritmo de clasificación basado en el conjunto de datos, el número de instancias y variables (características).

Algunas de las herramientas utilizadas fueron Microsoft Excel para el filtrado de datos y Google Colab, Jupyter, Python y SciKit-Learn para la implementación del modelo.

#### ii. Naives Bayes Gaussiano - Clasificación

Como lo expresa en [9] Naive Bayes es una técnica de clasificación estadística basada en el teorema de Bayes. Es uno de los algoritmos de Machine Learning más simple. Es un algoritmo rápido, preciso y fiable, inclusive tiene una alta precisión y velocidad en grandes conjuntos de datos. De igual manera en [11] formulan que Naive Bayes es uno de los más eficientes y efectivos algoritmos de aprendizaje inductivo para aprendizaje automático y procesamiento de datos. Su desempeño competitivo en clasificación es sorprendente.

Se utiliza una clase sklearn denominada SelectKBest, con la misma se procederá a tomar las cinco mejores características para proseguir utilizando las mismas en los distintos procedimientos a implementar en el modelo como se puede observar en la Fig 2.

```

X=dataframe.drop(['Aprueba'], axis=1)
y=dataframe['Aprueba']

best=SelectKBest(k=5)
X_new = best.fit_transform(X, y)
X_new.shape
selected = best.get_support(indices=True)
print(X.columns[selected])

Index(['Tiempo ', 'Avance', 'Realiza', 'Actividad', 'Participa'], dtype='object')

```

Fig 2: uso de la clase SelectKBest

Mediante el coeficiente de correlación de Pearson se puede comprobar de manera general la correlación de las características seleccionadas en el párrafo anterior dejando constancia una gráfica Fig 3, esta matriz es tomada como una métrica de análisis de datos con el fin de comprender la relación entre las diversas variables.

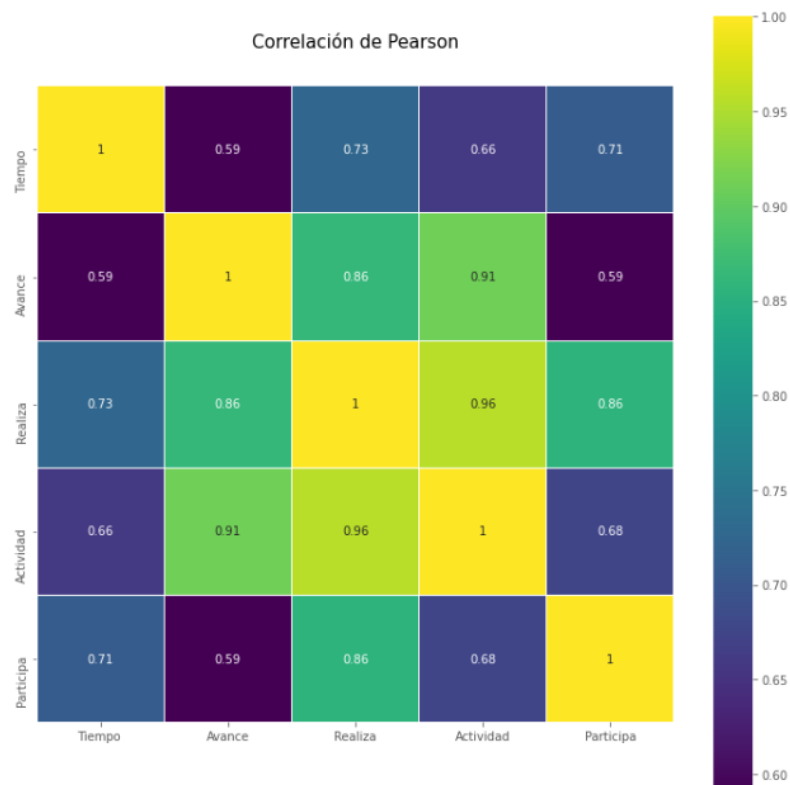


Fig 3: Correlación de Pearson del modelo.

#### 4. Resultados

Clasificar mediante aprendizaje automático a los estudiantes que tienen un rendimiento bajo en la utilización de la plataforma virtual del IES de Charata y permitir la implementación de estrategias alternativas de uso de plataformas virtuales o ecosistemas educativos multicomponentes y posibilitar la correcta toma de decisiones sobre los datos obtenidos.

Los resultados de la ejecución del modelo de clasificación de Naive Bayes Gaussiano aplicado al dataset “[iescharata\\_pru.csv](#)” se lo puede visualizar en el notebook Jupyter desarrollado en Google Colab y alojado en el [repositorio particular de GitHub](#).

Como se puede observar en la Fig. 4 se deja una captura de la funcionamiento del modelo con datos ingresados en Google Colab, observándose un la clasificación [1 0] en donde el 1 aprueba el taller y el 0 lo desaprueba.

```
[ ] # ['Tiempo', 'Avance', 'Realiza', 'Presenta', 'Foros']
    print(gnb.predict([[3.5, 72, 54.5, 29, 60],
                       [1, 25, 0, 15, 0]]))
    #Resultado esperado 0-Desaprueba, 1-Aprueba taller

[1 0]
```

Fig 4: resultado de la prueba del modelo.

#### 5. Conclusiones

Este trabajo presenta los resultados de una primera aproximación al machine learning y los modelos de clasificación para la toma de decisión basados en Naives Bayes y su aplicación a un conjunto de datos del ámbito la formación docente en la provincia del Chaco.

Se realizó un procesamiento de los datos mediante la creación de un dataset de prueba, con datos obtenidos del campus virtual E-ducativa del IES de Charata, se compartieron distintas visualizaciones y los link de todo el código en un repositorio de GitHub.

Naive Bayes como clasificador es muy utilizado en el procesamiento de lenguaje natural para la detección de spam, como así también en otras tareas a veces complejas, se puede utilizar para la detección de intrusiones en redes y como lo presentado en este artículo aplicaciones del ámbito educativo.

En el modelo presentado quedo demostrado que es rápido, simple de implementar, funciona bien con un conjunto pequeño de datos, se acomoda bien con muchas dimensiones y brinda resultados confiables sin cumplir las condiciones de distribución necesarias en el dataset.

## 6. Referencias

- [1] De-La-Hoz, E. J., De-La-Hoz, E. J., & Fontalvo, T. J. (2019). Metodología de aprendizaje automático para la clasificación y predicción de usuarios en ambientes virtuales de educación. *Información tecnológica*, 30(1), 247-254.
- [2] Esparza, G. G., Fuentes, M. D. L. M., del Real, T. A. R., & Reich, J. C. (2017). Un modelo basado en el Clasificador Naïve Bayes para la evaluación del desempeño docente. *RIED. Revista Iberoamericana de Educación a Distancia*, 20(2), 293-313.
- [3] Murphy, K. P. (2006). Naive bayes classifiers. *University of British Columbia*, 18(60), 1-8.
- [4] Rish, I. (2001, August). An empirical study of the naive Bayes classifier. In *IJCAI 2001 workshop on empirical methods in artificial intelligence* (Vol. 3, No. 22, pp. 41-46).
- [5] Nieto Acevedo, Y. V. (2020). UDLearn: modelo de aprendizaje de máquina que facilita la toma de decisiones académicas en las instituciones de Educación Superior.
- [6] Feng, Y., Zhou, M., & Tong, X. (2020). Imbalanced classification: an objective-oriented review. *arXiv preprint arXiv:2002.04592*.
- [7] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *the Journal of machine Learning research*, 12, 2825-2830.
- [8] Raschka, S., & Mirjalili, V. (2017). *Python Machine Learning: Machine Learning and Deep Learning with Python. Scikit-Learn, and TensorFlow*. Second edition ed.
- [9] Gonzalez, L. (2019). *Machine Learning con Python. Aprendizaje Supervisado*. Primera Edición. Maracaibo - Venezuela.
- [10] Osisanwo, F. Y., Akinsola, J. E. T., Awodele, O., Hinmikaiye, J. O., Olakanmi, O., & Akinjobi, J. (2017). Supervised machine learning algorithms: classification and comparison. *International Journal of Computer Trends and Technology (IJCTT)*, 48(3), 128-138.
- [11] Zhang, H. (2004). The optimality of naive Bayes. *AA*, 1(2), 3.