



UANL

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN



**NOMBRE:** WALTER RAUL PEREZ MACHINENA

**MATRICULA:** 1603647

**GRUPO:** 003

**NOMBRE ACTIVIDAD:** PRODUCTO INTEGRADOR DE APRENDIZAJE (PIA)

**NUMERO DE ACTIVIDAD:** 5

**MATERIA:** APRENDIZAJE AUTOMATICO

**DOCENTE:** JOSE ANASTACIO HERNANDEZ SALDAÑA



UANL

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN



FCFM

FACULTAD DE CIENCIAS FÍSICO MATEMÁTICAS

## Contenido

Instrucciones .....	3
Introducción .....	3
Hallazgos.....	4
Conclusiones.....	7
Referencias .....	7



# UANL

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN



## Instrucciones



JOSE ANASTACIO HERNANDEZ SALDAÑA 17/07 10:07 p. m.

### Requisitos PIA

utilizando los conjuntos de datos que adjunto, encuentra un modelo de clasificacion con sus parametros utilizando cross validation con el criterio ROC\_auc que te de un valor mayor de 0.75 en el conjunto de validacion y prueba. Incluye un reporte donde expongas los resultados y el codigo en el repositorio de git.

Adicionalmente, debido a un problema en los csv, por instrucciones del profesor solo se utilizará el train.csv.

### Objetivo:

- Encontrar un modelo de clasificación con sus parámetros, utilizando cross-validation con el criterio ROC\_auc

## Introducción

En esta actividad se utilizó un conjunto de datos que muestra información acerca de discursos de oratoria, la intención es encontrar el mejor modelo de clasificación utilizando el criterio ROC\_auc además de implementar una validación cruzada con el objetivo de obtener una puntuación mayor a 0.75. Para esto se utilizarán los datos proporcionados por el proceso identificados como "train.csv".

De igual manera se presentará de forma gráfica los datos obtenidos para una mejor interpretación de la información.



# UANL

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN

Hallazgos



Primero determinamos cual sería nuestra X y cual sería nuestra Y, considerando que es un modelo de clasificación se decidió tomar como variable Y a la variable “engagement”, y remover del conjunto de X a las variables “id” y “engagement”, la primera por considerarse irrelevante para el caso presentado y la segunda por ser la variable a obtener, es decir Y.

Modelo	ROC AUC Validación Cruzada
KNN	0. 783866
Decision Tree	0. 720124
Random Forest	0. 881452
MLP Classifier	0. 877614
AdaBoost Classifier	0. 866525
Gaussian Naive Bayes	0. 810303
Quadratic Discriminant Analysis	0. 837632

Con esta tabla podemos corroborar que el modelo con el ROC\_AUC y validación cruzada mas alto es el Random Forest, el cual tiene **0.881452**

De la misma forma se graficaron todas las curvas roc, en aquellos modelos que tuvieran una puntuación superior a 0.75

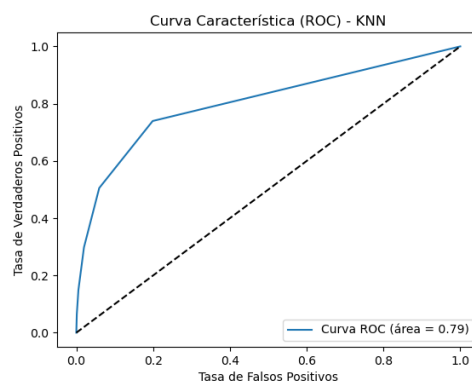
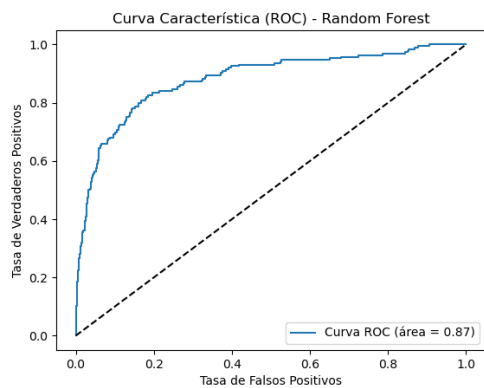
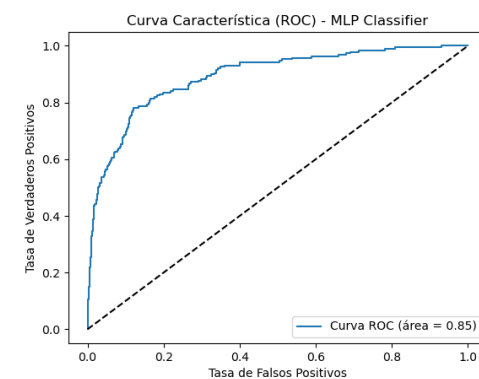
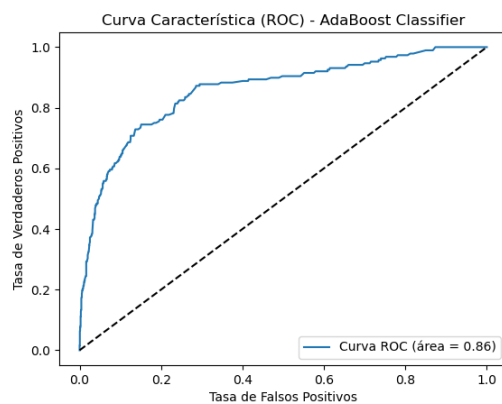
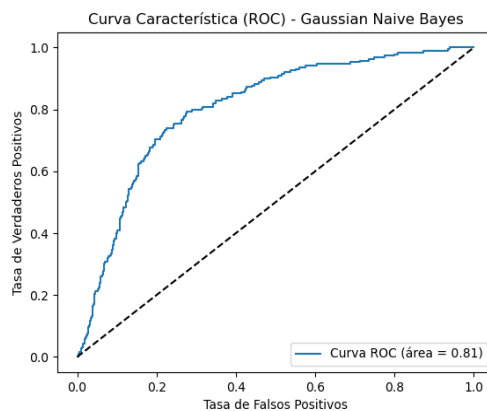
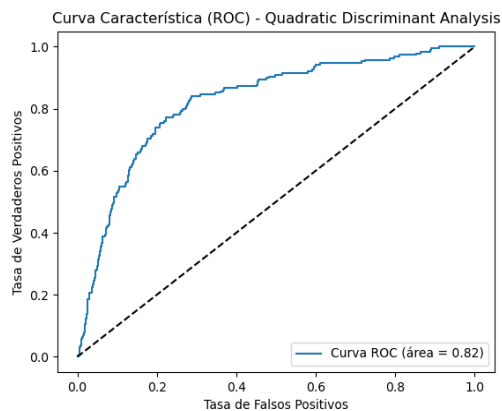


# UANL

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN

# FCFM

FACULTAD DE CIENCIAS FÍSICO MATEMÁTICAS





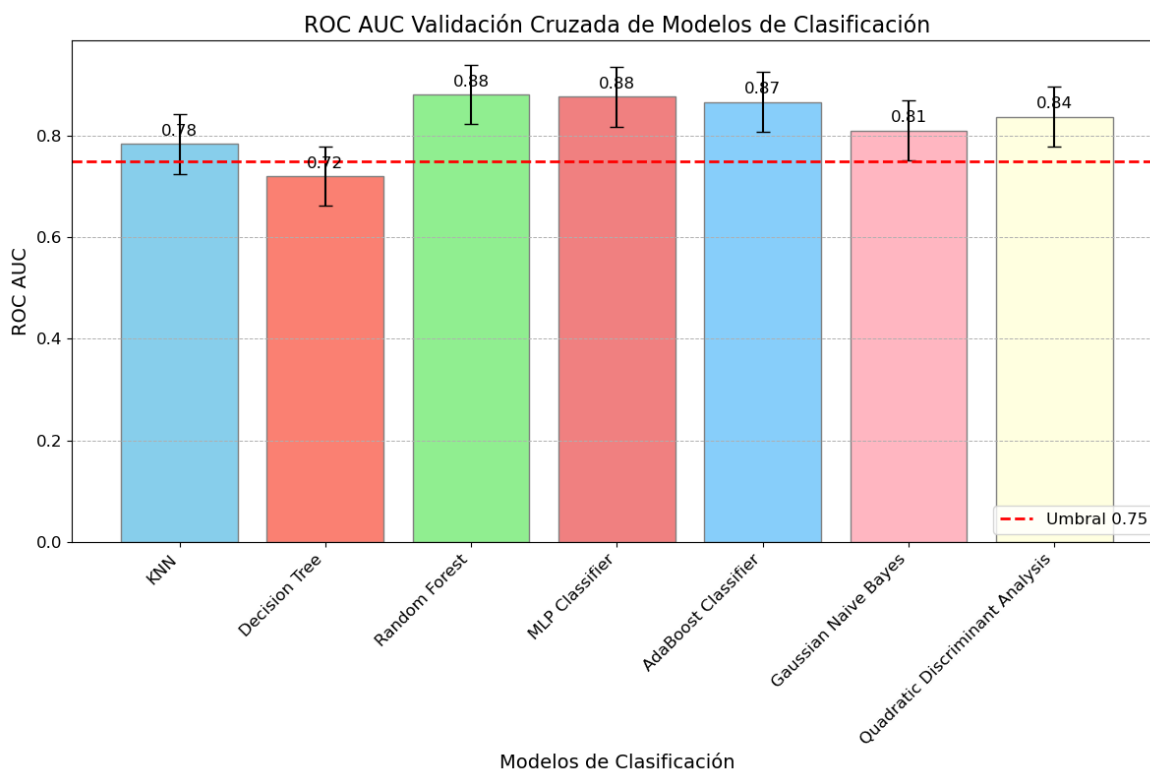
# UANL

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN



FACULTAD DE CIENCIAS FÍSICO MATEMÁTICAS

De la misma manera, se realizó una gráfica donde se muestran todos los modelos de clasificación, en la cual se muestra una línea que hace referencia al objetivo de 0.75. En donde se puede visualizar de forma más gráfica cuál es el mejor modelo donde se utilizó `roc_auc` y su validación cruzada.



De igual manera como en la tabla anterior, podemos ver que el mejor modelo **es el Random Forest**, estos modelos fueron tomados de la siguiente referencia: (Scikit-learn, 2024)



# UANL

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN



# FCFM

FACULTAD DE CIENCIAS FÍSICO MATEMÁTICAS

## Conclusiones

En la presente actividad se realizaron acciones de agrupamiento, donde se identificó el valor óptimo, siendo este el 2, esto se realizó de forma visual (gráfica de código – Gráfica 1) de esta manera se realizó un reentrenamiento ya con el valor óptimo identificado. También logramos visualizar la forma en que se realizó el agrupamiento de forma visual con la Gráfica 2.

En la presente actividad se identificó cuál era el mejor modelo de clasificación utilizando el criterio `roc_auc` y validación cruzada, siendo este **el Random Forest**. De la misma manera se graficaron todos los modelos que superaban el umbral de aceptación de 0.75, y se muestra gráficamente la puntuación obtenida para visualizar mejor que modelo es el correcto para este caso.

## Referencias

Guido, A. C. (2016). *Introduction to Learning with Python*. California: O'REILLY.

Scikit-Learn. (2024, 07 27). *Scikit-Learn*. Retrieved from <https://scikit-learn.org/stable/>

Scikit-learn. (2024, 07 27). *Scikit-learn.org*. Retrieved from [https://scikit-learn.org/stable/auto\\_examples/classification/plot\\_classifier\\_comparison.html](https://scikit-learn.org/stable/auto_examples/classification/plot_classifier_comparison.html)

Material de Clase.