

Prueba Teórica - Darwin Steven Berrío Álvarez

Cargo Especialista en Soluciones y Analítica Finanzas

1. Lo primero que debemos es entender muy bien es el problema que queremos solucionar y verificar las fuentes de información si son suficientes para abordar la solución requerida.

De acuerdo con la descripción del problema podemos utilizar un modelo de machine learning para la clasificación de los diferentes correos, a su vez contamos con datos históricos los cuales podemos preprocesar para adaptarlos y puedan ser consumidos por el modelo.

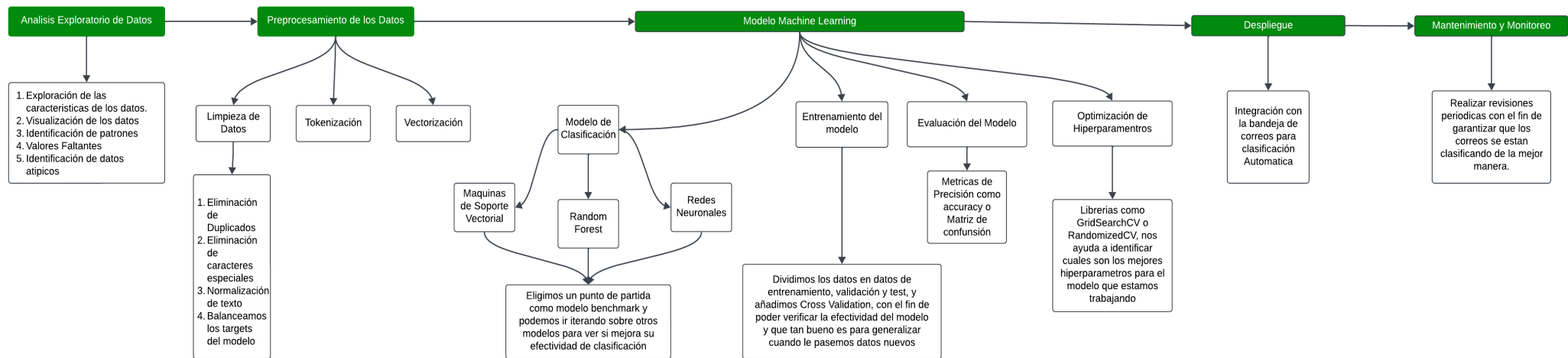
Para iniciar con este proyecto solicitaría a recursos humanos:

1. Una persona especializada en Ciencia de Datos
2. Una persona especializada en modelos de Procesamiento de Lenguaje Natural
3. Ingeniero en infraestructura en la nube
4. Analista de Datos

Y desde el área de Tecnología:

1. Herramientas para la implementación de los modelos como Python, Visual Studio Code, Github entre otros.
2. Plataforma de Azure para implementar los diferentes recursos que se necesita para poner el modelo en producción.

Arquitectura Modelo de Clasificación



2.

¿Cree que el modelo esté sufriendo Drift?

Los modelos de machine learning pueden sufrir de Drift cuando sus variables de entradas y su target cambia con el tiempo, si no hay un monitoreo constante que permita identificar la precisión y efectividad de los modelos, estos pueden sufrir cambios en sus variables que implica que con el tiempo estos modelos no sean viables para realizar las predicciones que necesitamos, estos cambios se pueden dar por múltiples razones como la manera en que se recoge la información del modelo, la interacción de los usuarios entre otros.

¿Cómo puede validarlo?

Podemos validar si nuestro modelo está sufriendo de Drift monitoreando las métricas de error, podemos realizar una comparación de que tan bueno esta siendo nuestro modelo y compararlo con las métricas que tuvimos al inicio si están han cambiado drásticamente este es un gran indicio de que nuestro modelo esta afectado por el Drift, también podemos validarlos a través de gráficos dispersión que muestren el comportamiento de los datos históricos y los datos recientes.

¿De ser así, que haría usted para corregir esto?

De ser que el modelo este sufriendo de Drift podríamos realizar diferentes formas que permitan corregir esto, tales como un reentrenamiento del modelo donde pueda incorporar datos nuevos para que el modelo pueda generalizar nuevamente de manera adecuada, otra forma podría ser tener una validación periódica del modelo para identificar que tanto sus predicciones están desmejorando y ya si los datos cambian de una manera constante deberíamos pensar en crear modelos adaptativos y más robustos que puedan integrar los cambios de manera más seguida, para que el modelo no pierda sus propiedades y este siga siendo efectivo a la hora de predecir mejor los valores.

3. Considero que lo primero que se debe hacer es estructurar bien el Prompt que se le realizaría al chat bot, de esta manera se le otorga información de contexto y precisa que permite al modelo obtener un marco de respuestas claras, de esta manera ayuda a mantener enfocado al chat bot.

Otra manera de mantener información clara y sin aleatoriedad considero que es la especialización del modelo, el cual se enfoque en un tema particular donde se le indica un limitante de conocimiento para dar sus respuestas e igualmente el modelo podría ser reentrenado con ejemplos de respuestas que deseemos y respuestas que no deseemos, o también podemos interactuar con parámetros que permiten controlar la creatividad y la coherencia de las respuestas del modelo.