**Especificación del modelo de ML**

1. Motivación del proyecto de ML

¿Cuál es el problema específico de ML que se quiere resolver?

Predecir el valor del dólar en pesos colombianos.

¿Cómo este problema se articula con el problema de analítica formulado?

El problema de analítica no está totalmente delimitado por lo tanto podemos decir que es el mismo problema que de maching learning el cual es predecir el valor del dólar en pesos pesos colombianos

1. Definición del problema

¿Cuál es la salida especifica que se va a predecir?

Nuestro objetivo es ver el valor del precio del dólar, siguiendo el orden temporal que se proponga ya sea en el entrenamiento o sea en los test. Podemos estar viendo que tal vez lo más simple pueda ser predecir si el valor sube o baja como en papers que ya han sido publicados.

¿Cuáles son los datos de entrada requeridos para el proceso de analítica?

Para el dataset a utilizar en ML describa:

1. Número de registros de entrenamiento:

Tenemos 10814 registros, pero no es posible determinar cuantos utilizaremos de entrenamiento, en efecto al final se utilizaran probablemente los 10813 mínimo además que cada día tendremos un dato mas

1. Existen datos no-balanceados, existen sesgos en los datos:

Tenemos datos totalmente confiables y limpios, estos son sacados del banco de la república.

1. ¿Qué tan frecuentemente se actualizan los datos?

Diariamente.

¿Cuáles son los factores más relevantes que permitirían predecir la salida?

Como solo tenemos dos datos no podemos hablar de demasiados factores, ahora como estamos hablando de una serie de tiempo los datos que más pesan son los datos últimos que tenemos.

1. Medidas de desempeño

¿Se tiene un marco de trabajo experimental de línea base?

Si:

¿Existe documentación?

Si

¿Cómo se construyó?

Este modelo utilizo los comentarios negativos y positivos que había en foros y discusiones sobre la economía turca y el cambio entre la moneda turca y el dólar. A partir de una puntuación que se le daban a los comentarios se categorizaban en positivos y negativos

¿Cuál es su desempeño y cómo se mide?

Se tuvo un 95.87 de accuracy con el modelo de HLT( Holt’s linear trend), además del accuracy también se utilizaron el mean absolute percentage error, mean absolute error y el mean squared error. Para medir la precisión de los modelos.

¿Cómo se va a medir el desempeño del modelo de ML?

Se van a utilizar estas dos medidas el accuracy y probablemente las más confiable sea el mean square error al no utilizar más variables de las que necesitemos.

¿Cuál son las implicaciones de tener FP y FN (errores)?

Que tendremos valores erróneos en nuestra predicción del valor del dólar, igualmente lo lejos que caiga del valor real nos dará mucha información sobre que tan cerca o lejos esta el modelo

¿Cuál es la literatura relacionada al problema (ej. papers y/o proyectos)?

En internet se encuentran varios papers relacionados, aunque siempre utilizan mas variables para determinar la relación con la tasa de cambio, en estos son diferentes monedas no el peso Colombiano.

Machine Learning-based USD/PKR Exchange Rate Forecasting Using Sentiment Analysis of Twitter Data. DOI:10.32604/cmc.2021.015872

<https://towardsdatascience.com/exchange-rate-prediction-machine-learning-with-5-regression-models-d7a3192531d>

US Dollar/Turkish Lira Exchange Rate Forecasting Model Based on Deep Learning Methodologies and Time Series Analysis

1. Partición de los datos

¿Qué consideraciones se deben tener en cuenta para la partición de los datos (sesgos, imbalance de clases, múltiples etiquetas)?

La consideración más fuerte es que nunca vamos a poder utilizar datos aleatorios los datos de prueba serán siempre posteriores a los datos de entrenamiento, esto para no alterar el orden temporal el cual es muy importante para las series de tiempo

¿Qué estrategia de partición se va a utilizar (random, LOOCV, K-folding)?

Train series partition folding machine learning

1. Línea de tiempo

¿Cuándo es el deadline?

23 de Agosto

¿Cuándo se deben los primeros resultados?

9 de Agosto

¿Cuándo se debe tener la solución propuesta?

13 de Agosto

1. Contactos

¿Quién es responsable del proyecto?

Juan David Peralta Rodriguez

¿Quién es el responsable del acceso a los datos (en training y test)?

Juan David Peralta Rodriguez

¿Quién puede ayudar a entender el proceso, los resultados (experto de dominio)?

Juan David Peralta Rodriguez

1. Recursos para entrenamiento

¿Cuáles son los recursos computacionales (software/hardware) disponibles para entrenamiento?

Dos computadores si es al caso utilizar un tercero

¿Cuál es el presupuesto (tiempo/dinero) disponible para entrenamiento?

Tenemos medio mes para lograr entrenar a la maquina con todos los desafíos que estén en medio

1. Despliegue

¿Cuáles son las restricciones tecnológicas para el despliegue del modelo (software particular, maquinas particulares)?

Vamos a tener un considerable gasto computacional entrenando a la maquina dado a que tendremos unas series cada vez más grande de datos para entrenar a la máquina y posteriormente testearla.

¿Cuál va a ser el uso del modelo (consultas diarias, mensuales, etc)?

La idea es que sea de consulta diaria.

¿Cuál va a ser el mecanismo de acceso al modelo?

Deberá ser un ejecutable el cual nos mostrara diariamente su predicción del valor del dólar para el día en cuestión

¿Cómo se va a integrar el modelo con otros componentes del negocio (personas otros sistemas)?

El sistema nos dirá diariamente si es conveniente hacer una compra en dólares, allí el gerente será el encargado de decidir si obedece a la maquina o decide utilizar su propio criterio para decidir si hacer una compra