Uso de XGBoost para la predicción del mercado financiero Forex.

Mawrer Amed Ramírez Martínez

ms723737@iteso.mx

ITESO

Asesor

Dr. J. Guadalupe Olascuaga Cabrera

jolascua@iteso.mx

ITESO

Dr. Luis Fernando Gutiérrez Preciado

lgutierrez@iteso.mx

ITESO

*Resumen*—La investigación de distintas técnicas para interpretar y predecir movimientos en el mercado de divisas interbancarias Forex es muy extensa, va desde aplicación de métodos de control, pasando por identificación de patrones en gráficas, hasta la aplicación de redes neuronales y distintos tipos de algoritmos de aprendizaje automatizado. Múltiples investigaciones muestran resultados positivos en la aplicación de algoritmos de aprendizaje automatizado. Es un problema complejo con gran cantidad de variables, se han desarrollado distintitos enfoques para atacar el problema como el llamado análisis técnico, análisis fundamental y combinaciones de estos. En este trabajo se aborda el problema desde un análisis técnico, donde se estudian las características principales de los datos históricos de la tupla interbancaria EUR/USD recolectados desde el 1 de enero de 2014 hasta el 30 de mayo de 2020, el autor realiza un etiquetado de clasificación en base al porcentaje de cambio al cierre de cada periodo, se realizan experimentos de optimización para identificar los mejores valores de clasificación; como parte del análisis de los datos de entrada se utilizan herramientas estadísticas como la autocorrelación y correlación entre distintas características. Para la selección de nuevas características se utiliza la librería Technical Analysis (AT) Library para obtener indicadores a partir de la serie de tiempo original, en base a modelo de tipo Clasificador XGBoost se obtiene un subconjunto de características relevantes que serán utilizadas por los modelos finales. Posteriormente se entrena el modelo clasificador XGBoost para finalmente utilizarlo para predecir un subconjunto de datos de prueba, los cuales serán evaluados mediante simulación de trading (backstesting) obteniendo una media de resultados favorables.

*Palabras Clave*— ForReign Exchange Market, Exchange Rates, Machine Learning, Long Short Term Memory (LSTM), XGBoost, Prediction.

# INTRODUCCIÓN

Este es un documento es una investigación del uso de modelos de aprendizaje automatizando para la predicción de movimientos en el mercado interbancario de Divisas FOREX.

En esta siguiente investigación se propone el uso del modelo de clasificación XGBoost para la predicción de movimientos de cierre en el mercado interbancario Forex,

Dentro de la investigación

(dejar en claro que el problema es complejo y que la cantidad de variables que afectan el tipo de cambio es muy grande, definir bien el tipo de problema), se …

Pasos del pipeline:

## Preparación de datos.

En esta fase los datasets iniciales de compra y venta fueron unidos en un solo archivo dando como resultado un archivo principal de once columnas [timestamp, Open\_ask, High\_ask, Low\_ask, Close\_ask, Open\_bid, High\_bid, Low\_bid, Close\_bid, Volume].

## Extracción de características

Se agregaron features como moving average,

Pct range

Diff pct range.

## Análisis y selección de características

Se utilizo el modelo de clasificación XGBoost de tipo

Con el cual a partir de un análisis de threasure se determinó usar un total de X características.

## Entrenamiento del modelo

Para el entrenamiento del modelo se utilizo un porcentaje de 20 para validación y .20 para pruebas.

## Validación del modelo

En la etapa de validación se obtuvo el acc, el ruc y el auc para determinar que medelo y con que parámetros es mejor.

## Evaluación

Para la evaluación del modelo se utilizó la librería de backtesting con una simulación de inversión inicial de 1000 usd.

Mediante las gráficas de velas es posible presentar los datos de entrada e identificar de forma simple movimientos positivos o negativos a través de la serie de tiempo.

Los datos utilizados de entrada utilizados son historiales del mercado Forex del banco Dukascopy [] del tipo de cambio Euro Dollar.

Para el desarrollo de esta investigación se utilizó el siguiente ambiente: Python 3.7.3. Con las siguientes librerías de y versiones: Backtesting-0.1.4 [], TA- 0.5.25 [], xgboost-1.1.1 [], keras- 2.3.1 [], tensorflow- 1.14.0 [].

# Trabajo Relacionado

Existen numerosas investigaciones relacionadas a la predicción de los mercados financieros. Algunas basadas en gráficas, las cuales intentan identificar patrones

(Hablar de análisis técnico, análisis fundamental, combinación, series de tiempo, características de series de tiempo, hablar de modelos y como hay evidencia que los mercados financieros están influenciados fuertemente por algoritmos).

# Análisis De Características De Datos

EUR/USD, se utilizaron frecuencias de 1M, 1H, 4H y 1D, de la ventana de tiempo 1 de enero de 2004 hasta el 30 de mayo de 2020.

Los datos de entrada contienen originalmente las siguientes columnas: *Timestamp*, *Open*, *High*, *Low*, *Close* y *Volume* tanto para el precio de compra como para el precio de venta en las distintas temporalidades.

A partir de los datos de entrada de generaron indicadores estáticos de tipo estocásticos o de movimiento promedio.

Con el uso de la librería AT fue posible obtener 133 características adicionales.

Los datos a utilizar son del tipo EUR/USD, fuente de datos. Recopilación y análisis de distintas frecuencias (1M, 1H, 4H, 1D). dando como resultado menor fluctuación y mayor ganancia en backstesting con el conjunto de datos de 1D.

Inputs del conjunto de datos principal: Time, Open, High, Low, Close y Volumen de tipo Compra y Venta.

Pct Change Diff, explicar autocorrelación.

Pct Change, explicar autocorrelación.

Hablar de los indicadores de tipo Osciladores, Moving Average, Estocásticos, Librería TA de indicadores técnicos.

# Algoritmos de Aprendizaje Automático

Como primer exploración se propuso predecir el precio de cierre con una

## Clasificación:

### Long Short Term Memory (LSTM)

### XGBoost (XGB)

### DNN

Explicación de cada uno de ellos, separación de subsets de datos, oversampling, k-Fold, randomizacion, cross-validation.

# Arquitectura de Modelos

Arquitecturas usadas en modelos de regresión, clasificación (LSTM, XGBoost, DNN)

Diagrama de todo el pipeline.

# Experimentos

## Experimentos de regresión (Close), (PCT Change) (PCT Change Diff).

## Experimentos de Clasificación

## Optimización de Backtesting.

En los experimentos realizados para backtesting se

## Experimentos de Acc y profundidad utilizada.

## Identificación de características relevantes con modelos XGBoost en distintos conjuntos de datos (1H, 1D) y distintas profundidades.

## 

# Evaluación De Modelos

Una vez el modelo XGBoost es entrenado es necesario evaluar

## Métricas de medición para evaluar el performance de los modelos con datos de pruebas.

## Uso de la librería backtesting para evaluación de resultados predichos.

# Resultados

Resultados obtenidos a partir de la evaluación de modelos con backtesting. Algo breve.

# Conclusiones

Mediante el uso del clasificador XGBoost fue posible

Agradecimientos

Agradecimientos CONACYT, etc**.**

Referencias

1. G. O. Young, “Synthetic structure of industrial plastics (Book style with paper title and editor),” in *Plastics*, 2nd ed. vol. 3, J. Peters, Ed. New York: McGraw-Hill, 1964, pp. 15–64.