ITESO - Maestría en Sistemas Computacionales

Asesor: J. Guadalupe Olascuaga Cabrera

Co asesor: Luis Fernando Gutiérrez Preciado

Alumno: Mawrer Amed Ramirez Martinez

Reporte de Avance de Trabajo de Obtención de Grado

Asesoría: 16 julio de 2020 – IDI4

**Completado:**

* Continúe las pruebas con LSTM 5C y 3C con frecuencias 1D y 1H. Los resultados en todas las pruebas fueron ‘menos peores’ (perdidas) con 1D. Por tal motivo se decidió continuar la fase de exploración con ese dataset como base.
* Agregar Logs para tracking de resultados.
  + Resultados por Modelo
  + Resultados por backtesting
* Optimización de valores para Backtesting de 3 Clases. 99% de Win Rate.
* Se exporaron los features de la librería backtesting para generar estrategias ganadoras usando Stop Loss (sl) y/o position.close().
* Se agrego un modelo de DNN básico 1 y 2 capas ocultas.
* Se implemento la librería TA para agregar features de indicadores financieros.
* Se agrego un modelo d XGBoost de tipo Logistic Regression para selección de Features.
  + Se encontró resultados superiores a los alcanzados con LSTM actual. > .45 acc
  + Se encontró que los resultados son mejores utilizando 3 Clases vs 5 clases.
  + Para la selección de features se exploró con datasets multiples de 27 features, 70 features y 145 features.
    - En base a estos resultados se exploró el .acc utilizando n de profundidad para los 3 casos y se compararon resultados. Todo para frecuencia de 1D
  + Los mejores resultados en ACC de XGBoost animaron mi exploración de mas modelos con esta herramienta.
    - Se identifico que a mayor distribución de la clase es mejor el accuracy, se logro acc > a .87 con valores con mayor margen. EL problema fue la cantidad de posiciones que se abren (predichas) es mas reducido.
    - Un problema extra es la distribución de las clases no están balanceadas.
    - Con datos balanceados, se tiene un accuracy menor del .30 pero mayor cantidad de posiciones -1 y 1.
    - Con los valores encontrados en la optimización no fue el mejor accuracy en el modelo.
  + Se realizo análisis de profundidad para multiples datasets (1D, 4H) con 145 Features, Selección de Features ~130F y 70F.
    - Mejores resultados promedio: 0 de Profundidad. 4, 8 y 11.
  + Se obtuvieron resultados ganadores hasta de un 30% de return en capital usando datos de prueba, nunca vistos por el modelo.
    - Con XGBoost Principalmente.
    - La mayoría de pruebas realizadas 1D
      * Los Dataset de 70 Features 145 Features
      * Profundidad: 0, 4 principalmente. Tambien 6, 9, 12, 7, 5 y 11
      * .002 de validación y .2 de pruebas.
    - El Accuracy de los modelos es de mas de .39
    - Hay casos donde el accuracy es superior al .50 y no se tienen ganancias.
    - Backtesting Model Name:
      * 3C-EURUSD\_FULL\_1D\_145F\_0.002Sep\_\_5137Size\_XGBOOST\_0Prof\_0.39acc\_Class\_02\_RealTest
* Draft del Articulo con las subsecciones.

**Pendientes Agregar:**

* **Continuar exploracion teniendo en cuenta la probabilidad en las predicciones.**
* **Agregar Oversampling con las nuevas clases. Para DNN con2D. DONE**
* **Hacer un Shuffle para confirmar que el 30% es consistente independiente del pedazo de la serie de tiempo.**
* **Agregar mas metricas, no de ACC.**

|  |
| --- |
| EL dataset con todos los eventos hacia atras. |
| Unir el restante teniendo en cuenta la profundidad perdida |
| Balanceo - DONE |
| Hacer Shuffle - DONE |
| Fit |
| Backtesting con el otro.- DONE |

**Usados por el prof:**

|  |
| --- |
| .82 acc en validacion |
| 20 eventos hacia atras |

**Metricas que valen la pena.**

|  |
| --- |
| Categorical ACC |
| RUC |

**Agregar 24 Features nuevos:**

'EMA\_Close\_5','Momentum\_1', 'Momentum\_2', 'Momentum\_5', 'Momentum\_24', 'rsi','Rolling\_Mean25', 'Bollinger\_High', 'Bollinger\_Low', 'PP', 'R1', 'S1','R2', 'S2', 'R3', 'S3', 'SO%k', 'SO%k1', 'SO%d2', 'Trix\_5', 'MACD\_5\_25','MACDsign\_5\_25', 'MACDdiff\_5\_25', 'Mass\_Index\_25\_8\_9'

**Optimizar Backtesting a partir del etiquetado actual.**

**Obtener mejores valores para crear Clases. DONE**

**Eliminar Posiciones de Perdida y Re etiquetar.**

**98 % Win Rathe**

**Implementar Red Neuronal 2 Dimensiones con Balanceo**

**Findings:**

* **Etiquetado mas separado mejora el acc: hasta 86%, la mayoría de predicciones son 0, eso ayuda en la estrategia, y las posiciones que se buscan son**
* **1D 145F XGBoost la profundidad de 8 fue la mejor .539 acc. .0029 y -00.29 de Clasificacion.**
* **Hasta cerda de 30% de return, con 1D y 4prof con selección de Features**

**Tengo un return % de 30%. Identificar con que profundidad de realizo y sobre ese cerrar.**

**Mejorar estrategia de Trading**

**Como cerrar una opracion con stop loss - DONE**

**Generar documento para hacer backtesting de las pruebas.**

**Trabajar con probabilidades de Y en backtesting**

**Probar con KFold**

**Agregar SVM**

**Agregar mas Features y hacer XBoost**

<https://colab.research.google.com/drive/1eWNmtIlnjjnQmbFAF9DWb0UgBZx4oHJj#scrollTo=h14oBXAtvUYE>

**A screenshot of a cell phone

Description automatically generated**

**A close up of a building

Description automatically generated**

<https://machinelearningmastery.com/feature-importance-and-feature-selection-with-xgboost-in-python/#:~:text=Feature%20Selection%20with%20XGBoost%20Feature,feature%20selection%20in%20scikit%2Dlearn.&text=This%20threshold%20is%20used%20when,dataset%20and%20the%20test%20dataset.>

**~~Agregar Graficas de Loss para confirmar no este overfiteado~~**

**~~Que abra transacción una por una.~~**

Fuentes:

[1]

“Reducción de la pérdida: Descenso de gradiente estocástico,” *Google Developers*, 2020. [Online]. Available: https://developers.google.com/machine-learning/crash-course/reducing-loss/stochastic-gradient-descent?hl=es-419. [Accessed: 18-Jun-2020].

‌

Oversampling OSTSC. LSTM

<https://cran.r-project.org/web/packages/OSTSC/vignettes/Over_Sampling_for_Time_Series_Classification.pdf>

[1]

Using XGBoost For Feature Selection, *Kaggle.com*, 09-Sep-2016. [Online]. Available: https://www.kaggle.com/mshih2/using-xgboost-for-feature-selection. [Accessed: 11-Jul-2020].

‌

Notas PASADAS:

Metricas

AUC para clasificación, derivación del ROC. Reciving operator clasifier

RUC

Matriz de confusión.

Balancear sus datasets. Oversampling QAV, Trades , TBBAV, TBQAV, DownTrade, Uptrade, PctSum, PCT acumulado, MV19, Upper, Lower, Position.

Técnicas de balanceo.