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: 09 julio de 2020 – IDI4

**Completado:**

* Pipeline End to End para clasificacion de 5 clases (18 Features Actuales).
  + .24 Acc en datos de 1H
  + Metricas Acc and Recall
  + Analisis con Matriz de Confusion
  + CrossValidation
  + KFold (10 iteraciones, randomize).
* Se agrego OneHot para el Vector Y basado en el % de cambio.
* Prediccion multiStep para Regression con LSTM (dinamico a 2 o n steps).
* 14 modelos generados durante pruebas.

**Pendientes Agregar:**

* Random Oversampling Imbalanced Datasets (No possible para mas de 2 dimensiones)
* Una funcion de Logger con resultados, paths y características de los modelos. - DONE
* SVR SVM
* Selección de Features (Clustering)
* Agregar NASDAQ, S&P 500, Dow 30, MACD,

**TODO**

**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'

**Generar Dataset de 3 clases 45 Features - DONE**

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

**Obtener mejores valores para crear Clases.**

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

**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.