SICONG ZHAO PORTFOLIO

Sicong is currently a Master's candidate of Data Science at Duke University, and he is actively looking for a data science position.

[Download CV, Send Me an Email]

Predicting Emotional States Using Wearable Devices

This project is about determining if emotional states can be predicted from wearable activity sensor data. In order to achieve our goal, I have built models which classifies emotion states using data collected from the Motivated Cognition and Aging Brain lab, which contains:
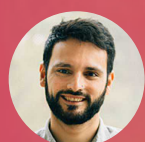
• measures of personality and behavior
• demographic data
• physical health metrics
• activity tracking data (Fitbit)
• functional brain connectivity

Currently, our best performance model is the one with CatBoost, with an accuracy of 49.5% (5-class classification) for predicting positive emotion score.

Teammates

Mikella Green

Joaquin Menendez

Predicting Emotional States Using Wearable Devices

MY CONTRIBUTION:

So far, I am in charge of feature engineering and modeling.

I have created 42 meaningful features from band data (steps & heart rate by minute), including resting time, activity level and variation of heart rate. Among all these features, the 'variation of heart rate in last 30 mins' performs the best, its importance score ranks 10th among 86 features.

I have also tackled this problem using XGBoost, CatBoost, and Linear Regression. CatBoost performs the best, because when compared with XGBoost, it accepts categorical data, and the best predictor is 'Marital Status', which is categorical.

TECH I UESD

- CatBoost
- XGBoost
- Linear Regression

The code is written in Python, and you can find them here in Google Colab.

NEXT STEPS

- Try Statistical Approach: Hierarchical Model
- Include brain connectivity data, and then redo the analysis. (I have not get the data yet)
- Try Deep Neural Network (Based on the size of our dataset, I do not think this is promising, but I would like to try it if time permit.)

Value Investing In Python

This is a tutorial inspired by Preston Pysh.

I learned his video tutorial about fundamental analysis for stocks, and then researched data source and implemented a python program to automatically conduct fundamental analysis for all 6000+ stocks in us market.

I have posted my tutorial on Medium, and then been invited to publish my articles on 2 Medium Publications: The Capital & Analytics Vidhya.

The tutorials:
1. Syllabus
2. Collecting financial data for fundamental analysis
3. How to Generate these Popular Stock Terms using Python
4. How to calculate the intrinsic value

TECH I UESD

• Linear Regression

• Selenium

• Pandas, including `unstack()` and `pivot()` function

The code is written in Python, and you can find the intrinsic value calculation part here in Google Colab.
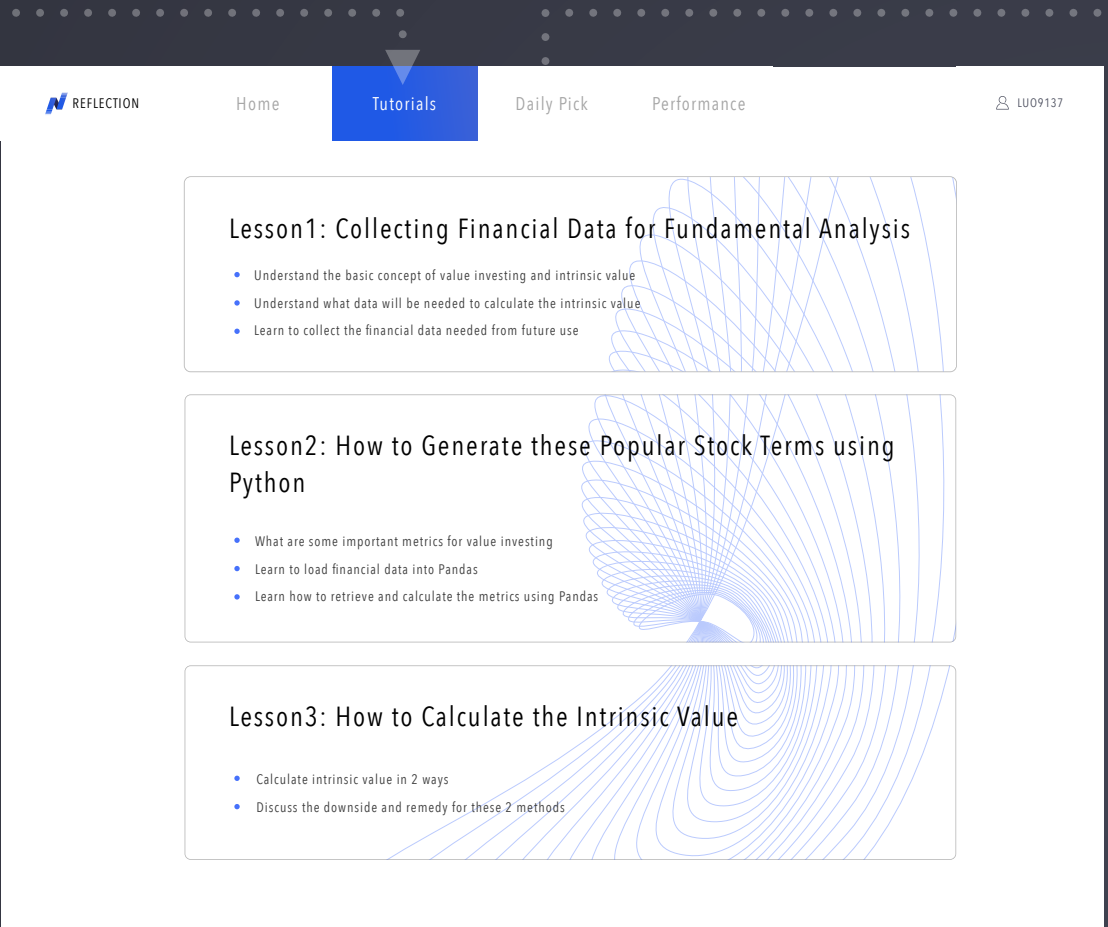
NEXT STEPS

• Automate the intrinsic value calculation

• Serve the results in a webpage. The design is in next page.

Value Investing In Python

Intro page
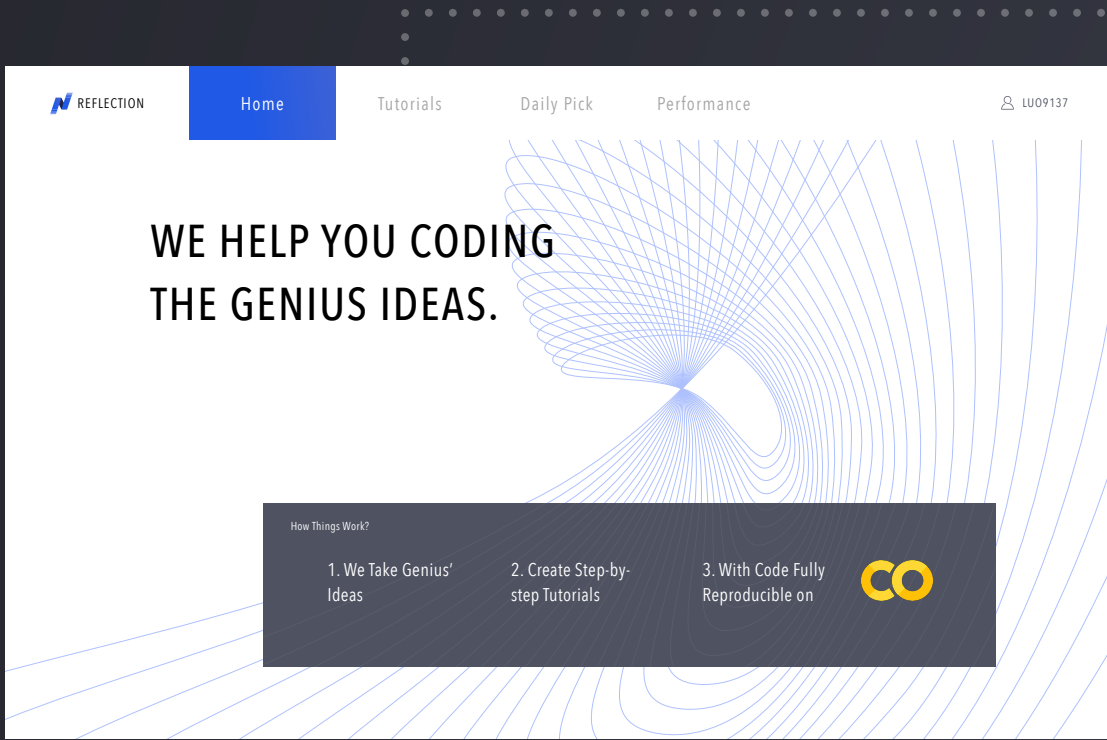
Tutorial Page

This page lists all our tutorials.

Daily Recommendation

We will process the stock market data on daily basis and recommend stocks based on the criteria in our tutorials.

Performance Tracking

We also track the performance of daily recommendation. Just so we can figure out the quality of strategies in the long-run.

Sorry for the space. But this slide is huge, please zoom in for more detail.

Cardiology Decompensation at Duke University Hospital

OVERVIEW

The end goal of this project is to help doctors detect risks of cardiology decompensation as early as possible. We collaborate with Duke University Hospital (DUH) to develop a dashboard that indicates cardiac decompensation risks for each patients.

MY CONTRIBUTION

- Extracting data points from historical database
- Help define phenotype definition (rules that generate outcome) from a data science's perspective
- Implement algorithms that determine if an encounter meets any phenotype definitions
- Calculate mortality rate for encounters with different types of phenotype combination

TECH I USED

- SQL
- Pandas / Numpy
- Tableau

NEXT STEP
1. Use CatBoost and RNN to build a predictive model
2. Integrate Dashboard