# Private Equity: Market Indicator Analysis

**Project Report for IEOR 4742 (Group 10)**

**Group Members:**

1.Nikhil Kamoji (nk2858)

2. Michael Pelton (mdp2176)

3.Susan Chen (zc2432)

4. Ziyu Zhang (zz2581)

# EXECUTIVE SUMMARY

# OBJECTIVE

The objective of this project is to identify at any point in time, the relative investment potential and attractiveness of certain sectors (for example: healthcare, energy etc.) for a private equity fund given the current market and economic environment. The end goal is to build a deep learning model that will rank major market sectors for growth potential and investment attractiveness that may eventually be used by a PE fund to guide sector targeting at a particular point in time.

As an example, a question that the results of the model might answer could be something along these lines: “Is energy a good bet right now given the fund horizon / market predictions etc.?”

# ITERATIVE JOURNEY AND KEY ACTION ITEMS

This section is meant to give the reader a high-level overview of the iterative journey that our team took over the course of the semester. It involves an overview of what we set out to build, how we thought at key junctures and an overlay of the model.

|  |  |
| --- | --- |
| Stage | Thought Process & Action |
| Initial interpretation of the problem | * To build a deep learning model to rank sectors as in the order of their investment attractiveness for a Private Equity (PE) fund. |
| Initial research methodology and findings | * Our first task was to find relevant data. Analyzed data from Bloomberg, CapitalIQ, Pitchbook and other public online resources. * Realized that using public market data as a proxy for private markets was the best way to move forward. Two reasons – Lack of sources that went back far enough in time & non-availability of ‘sector-wise’ private market data. |
| Narrowed down on dataset | * Sector SPDR (ETF) data set for selected sectors. |
| Extension of the available data | * Extended, cleaned and shifted the data to make it appropriate for our model (Detailed in latter sections) |
| First build of the model | * We played around with a few standard deep learning models to understand what best fit our purpose. |
| Built the model | * Built two versions, a feed forward model and an LSTM model |
| Iterations | * Went to several iterations and modifications on the data and retrained several times to achieve the best results |

# DATASET

The dataset used is a sector SPDR ETF dataset. We were able to source a 10-year dataset divided by sector. The following 9 sectors were used:

* Consumer Staples - includes companies that produce items such as food, beverages and non-durable household and personal products
* Consumer Discretionary - goods and services that are considered non-essential by consumers, but desirable if their available income is sufficient to purchase them
* Technology - The technology sector is made up of companies that, among other things, manufacture consumer electronics and their components, develop software, and provide information technology (IT) services like cloud hosting
* Health Care - Companies that provide medical services, manufacture medical equipment or drugs, provide medical insurance, or otherwise facilitate the provision of healthcare to patients.
* Industrials -
* Utilities -
* Materials -
* Energy -
* Financial Services -

We have 9 main data sets in csv format named for e.g. ‘data\_XLB’ etc. Each dataset has the following data points:

1. 20-year daily returns of the XLB sector (in this example)
2. 20-year daily returns of all the other 8 sectors
3. 20-year daily values of the following macro-economic indicators:
   1. Gross Domestic Product (GDP)
   2. LIBOR
   3. Mortgage Rate
   4. Real User Cost
   5. Unemployment Rate

# DATA PREPARATION

To prepare our data for model ingestion and the eventual output of sector rankings, the data sets had to be ingested from multiple sources and in multiple formats also at different rates of data collection, daily, quarterly and monthly. In order to do this, 4 notebooks were created that referenced both raw data and the outputs from previous notebooks. The 4 notebooks are as follows

combined\_indicators.ipynb – read in and process macroeconomic indicator data. Data comes in at differing intervals from daily to quarterly so unique dates must be identified, data must be read in and then concationated as to populate one dataframe for output into the next step. Output dataframe is indexed by date and each column represents an indicator. Output of this notebook is ‘combined\_indicators.csv’

combined\_sector.ipynb – read in and process sector data for all 9 sectors in scope for our model to rank. Output is a dataframe where rows are dates and columns represent each sector. Output of this notebook is ‘raw-data.csv’

daily\_data\_minipulate.ipynb – combined\_indicators.csv and raw-data.csv are taken as inputs here. Dataframes are joined and data is forward filled in order to account for some indicator data coming in monthly and quarterly. Output of this notebook is daily\_final.csv

Data\_model\_shift.ipynb: shifts daily\_final.csv data in a way as to allow for model to predict movement 125 days into the future and eventually let us recommend sectors for investment. This time frame of 125 days can be changed by the degree of shift. Output of this notebook will be shifted data for the models to be trained on for each individual sector. Data saved in the data folder.

From here we have model data to train on for each of the two types of models run. Then future shifted data to use for forecasting a half year into the future, data and future folders respectively contain these two types of sets.

# MODEL AND PARAMETERS

* Goal of the model: Rank the sectors based on percentage change in returns on predicted return values.
* Models for this practice were run on data shifted to forecast 125 days into the future, this can be adjusted to 4-5 years by changing portions of the data preparation notebooks above.
* We use the returns of all other sectors along with macroeconomic indicators to predict the returns of any given sector.
* We used two deep learning models on the data and compared the performance of both:

1. Deep Feed-forward neural network
2. Multivariate Time Series Forecasting Using an LSTM network
3. Deep Feed-forward neural network: project-DNN.ipynb

Model is run for a specified sector and input will be the shifted sector data described above.

With 20 epochs

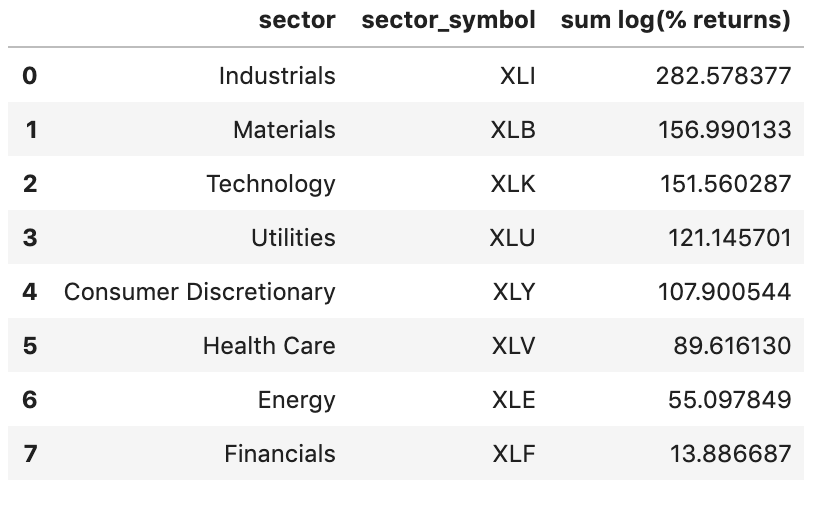
1. LSTM: project\_lstm.ipynb

Model output ingestion and ranking:

The output code packages the model and runs with input data sets and future sets for forecasting for each of the in scope sectors. This process occurs independently within each of the notebooks outlined above for the two types of models. For each type of model, the code is run for each sector to produce an output forecast of log daily returns for the next 125 days. This output is housed in a list where each position in the list is a day with log daily returns. The following functions combined returns together for each sector and rank the sectors by by forecasted potential returns. The end output for each of the two types of models can be found below.

# END MODEL OUTPUTS

Deep Feed-forward neural network:



LSTM:

