PURDUE UNIVERSITY | The Data Mine

Members: Isslam Alhasan, Ali DeLuca, Mihika Ghaisas, Sankaran Iyer, Kyochul Jang, Kris Leungwattanakij, Michelle Lin, Annie Zheng
TA: Oishi Ghosh

ABSTRACT

The objective of this project is to build a predictive model that helps UPS identify potential companies that will yield positive results for the company. To accomplish this, our team took on the responsibilities to perform research by reading books outlining the complexities of private equity firms, looking into individual companies to learn more about them, and reading through multiple articles about different types of machine learning models. In preparation of building the model, we cross referenced and condensed the data set and selected certain variables to use for the model. We built and tested several (supervised and unsupervised) models to determine which ones gave the most accurate results. Analysis of the model results indicated that the random forest model gave us the best result of 83% accuracy. Moreover, we were able to identify the correlations between companies based on the unsupervised models we created, namely KMeans++ and Dendrograms. Ultimately, to display results of our work, we decided to create a web app that better connects the user to our predictive model. Our future goals aim at improving the data, exploring more complex models and creating a web app containing more functionalities.

BACKGROUND

Important Terms and Definitions

Private Equity: a form of private financing involving investors who directly invest in companies or engage in buyouts of companies

Portfolio Company: a public or private company that private equity companies hold an interest in

Investment Process: process by which private equity companies decide to work with, then officially partner with portfolio companies (Figure 1.1)

Portfolio Company Valuation: how private equity companies determine the value of a portfolio company (Figure 1.2)

FINANCIALS
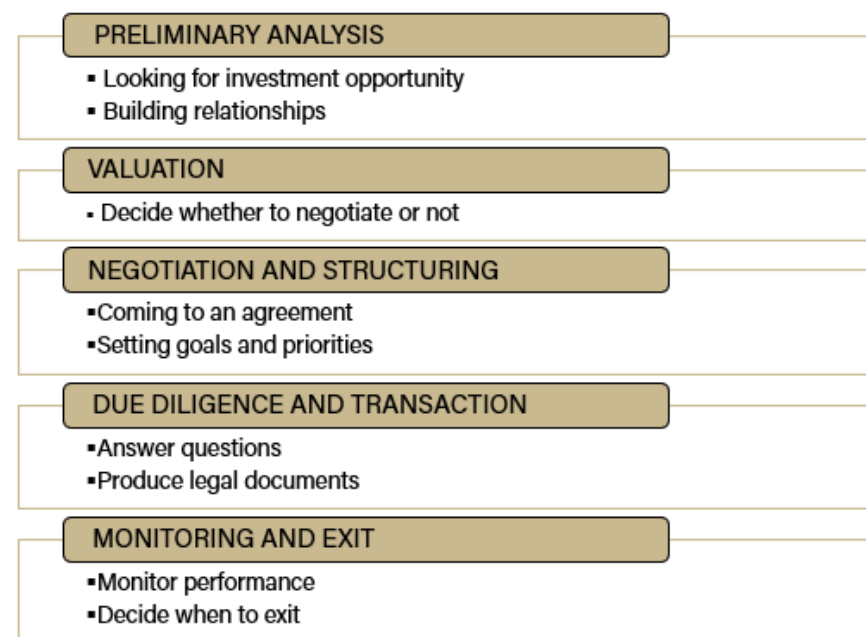- Looking at factors like IRR and TVPI to determine financial standings and future of a portfolio company

MANAGEMENT
- Observing the managerial styles to determine it they are working and if they will be successful in the future

TECHNICAL EXPERTISE
- Deciding if a given company's employees have the expertise to be successful now and, in the future

Figure 1.2 - Terms

Literature Review: Investment Process

PRELIMINARY ANALYSIS
- Looking for investment opportunity
- Building relationships

VALUATION
- Decide whether to negotiate or not

NEGOTIATION AND STRUCTURING
- Coming to an agreement
- Setting goals and priorities

DUE DILIGENCE AND TRANSACTION
- Answer questions
- Produce legal documents

MONITORING AND EXIT
- Monitor performance
- Decide when to exit

Figure 1.1 - Investment

Definitions of Variables Used

Companies: portfolio companies, as defined above

PE Firms: private equity firms, as defined above

HQ City: city location of portfolio company's head quarters

HQ State: state location of portfolio company's head quarters

HQ Location: city and state location of portfolio company's head quarters

Primary Industry Group: groups companies based on common business lines such as Transportation, Healthcare Services, and Software

Primary Industry Sector: sections of the economy classified by their purpose, such as Consumer Products and Services, Business Products and Services, and Healthcare

Revenue: revenue generated outside of that generate with UPS partnership

Net Amount for years 2020 and 2021: revenue generated in partnership with UPS in the years 2020 and 2021

Volume: average volume of packages delivered through UPS partnership

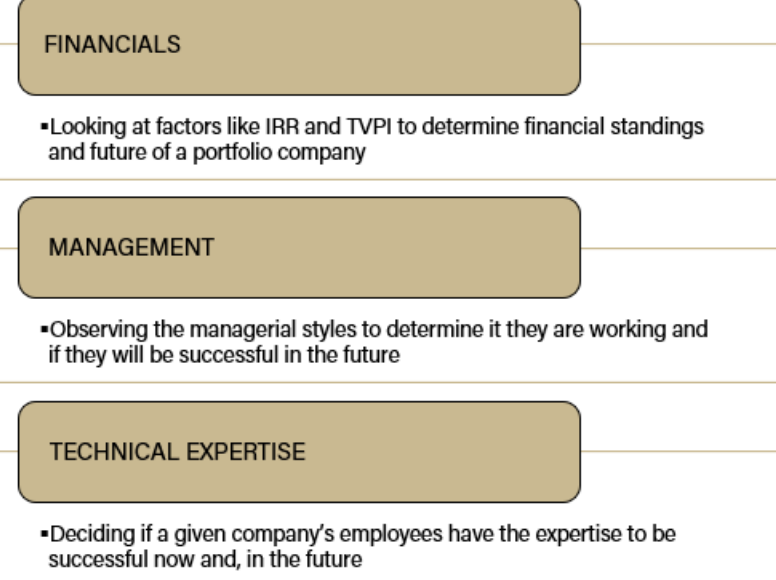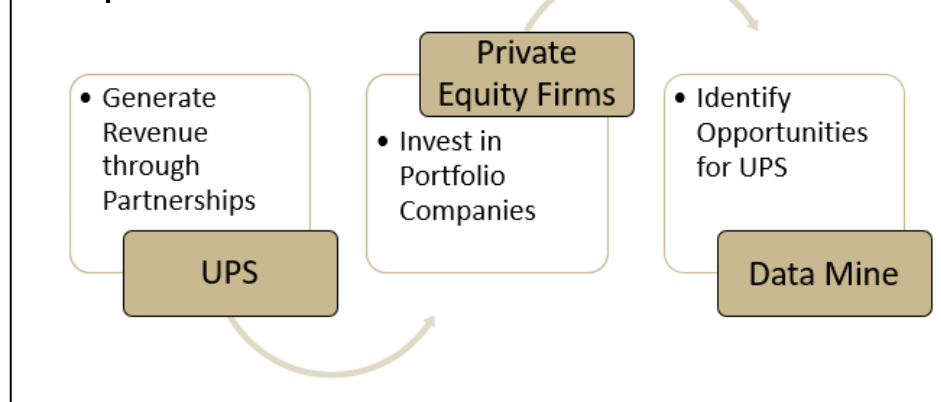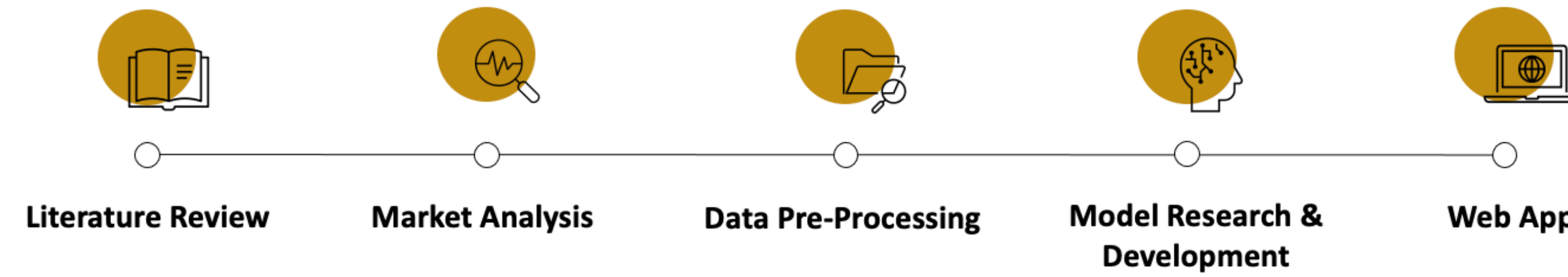**Weight:** average weight of packages delivered through UPS partnership

PROJECT GOALS

- Conduct research and analytics on PE Firms and Portfolio Companies
- Analyze variables that impact revenue generation for UPS
- Cross-Reference internal and external data
- Create and develop a predictive model to identify potential opportunities (portfolio companies) for UPS to partner with.

PROGRESS TIMELINE

Literature Review → Market Analysis → Data Pre-Processing → Model Research & Development → Web App

DATA PRE-PROCESSING

Data Analysis and Pre-Processing
- Understand Variables, Data Types, and Data Format
- Analyze & Research Patterns and Analytics to gain insights on variables that are key influencers
- Formulate, Cleanse, and Prepare Data for Analysis
- Feed Data into Model (Conversion of Categorical values to Numerical Type)
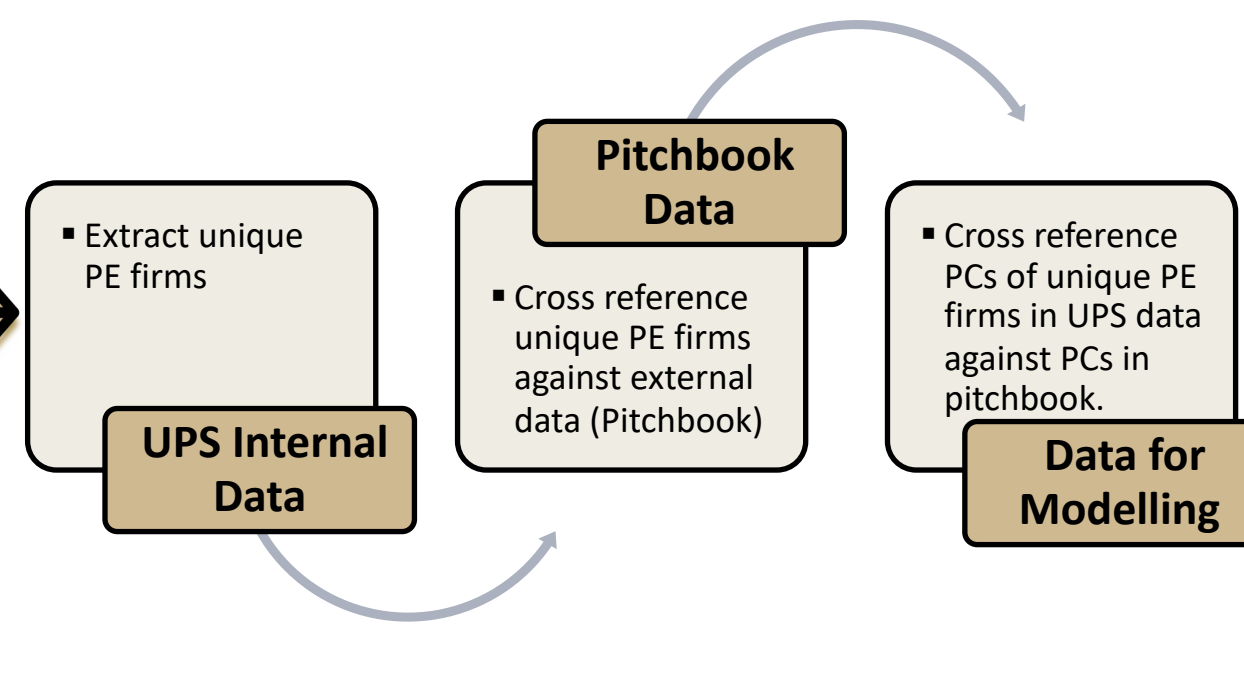
Cross-Referencing Data
- Extract unique PE firms
- UPS Internal Data
- Cross reference unique PE firms against external data (Pitchbook)
- Pitchbook Data
- Cross reference PCs of unique PE firms in UPS data against PCs in pitchbook.
- Data for Modelling

Figure 2.1 – Data Cleansing Process

MODEL RESEARCH & DEVELOPMENT

Why We Chose K-Means Model

Using the UPS dataset which consisted of Portfolio Companies, we wanted to categorize the companies in groups based on similarity. The K-means clustering algorithm finds groups which have not been explicitly labeled in the data. This can be used to confirm assumptions about what types of groups exist. (Figure 2.2)

Unsupervised Learning
Unsupervised Learning uses machine learning to identify patterns, similarities, and clusters within a dataset. It does not provide an output, but the goal is to learn more about the data distribution → **KMeans Model**

Supervised Learning
Supervised Learning uses machine learning to train the data and predict a result by allowing the model to learn the patterns between the data → **Random Forest Model**

```
In [17]: from sklearn.metrics import r2_score
         print("Accuracy Value (r^2):")
         r2_score(test_labels, predictions)

Accuracy Value (r^2):
Out[17]: 0.82804918079623341
```

Figure 2.2 – Predictive Models

Why We Chose Random Forest Model?

Our end goal is to predict the revenue a new portfolio company would be able to generate for UPS through their partnership. The Random Forest algorithm constructs a multitude of decision trees at training time and outputs the mean/median values of the trees, allowing it to predict the revenue generated for UPS. (Figure 2.2)

Figure 2.6 – Model Results

Model Research

After researching many different models, we chose the KMeans and Randon Forest models based on consideration of variables such as accuracy, ease of use, dataset size, and training times. We used Kmeans to create clusters of the data and confirm our market analysis and variable research. This helped us create our supervised model which predicted how much revenue UPS generated through partnerships. Some of the test results between the predictions and actual values aare displayed above(See figure 2.6).

Dendrogram Interpretation

The dendrogram basically shows how close companies within a cluster are related to each other. In the example above, Company 0 and 2 are under the same family and so are Company 3 and 4 which means they are related. However, the height of Company 0 and 2 family is less, indicating a strong relationship between them. (See figure 2.7)

Figure 2.7 – Dendrogram Interpretation

MARKET ANALYSIS

- 4 Companies contribute for 84.52% of the Net Amount Revenue UPS Generated
- Researching those PE Firms indicated importance of company size, and industry group/sector
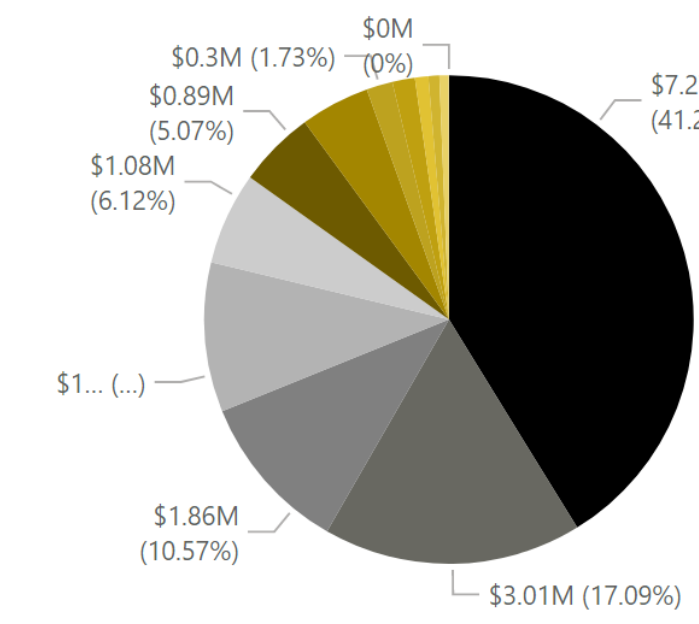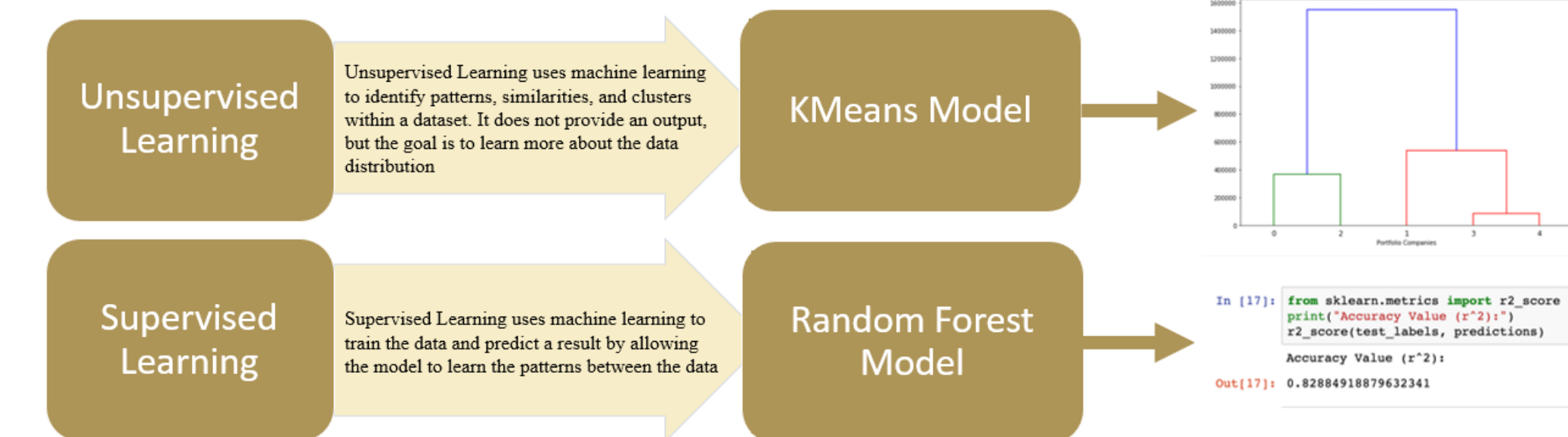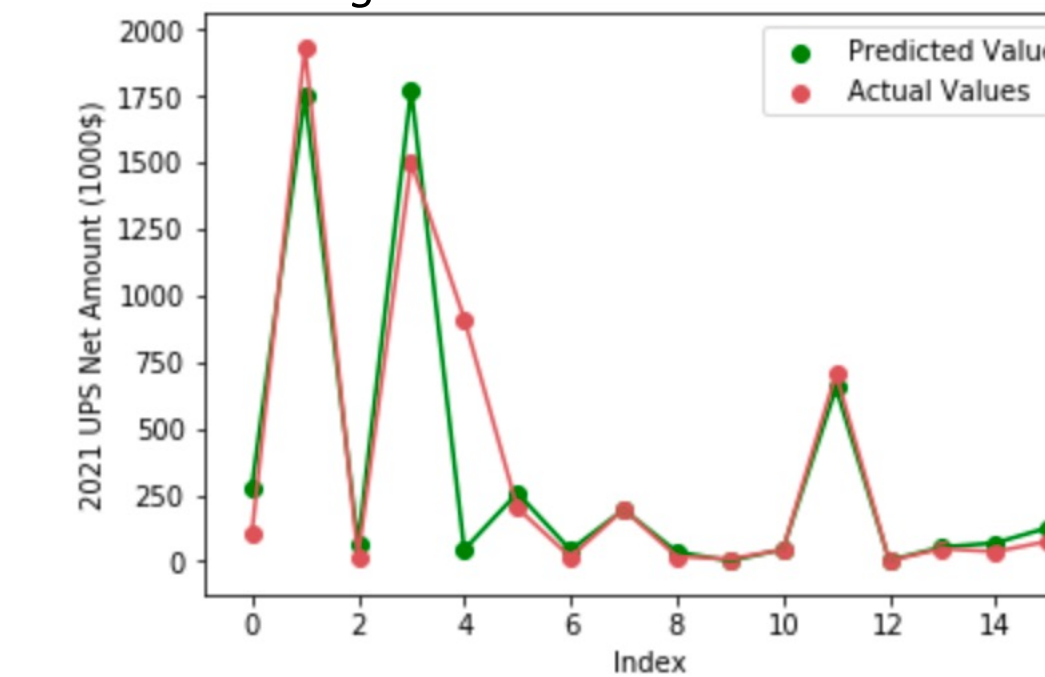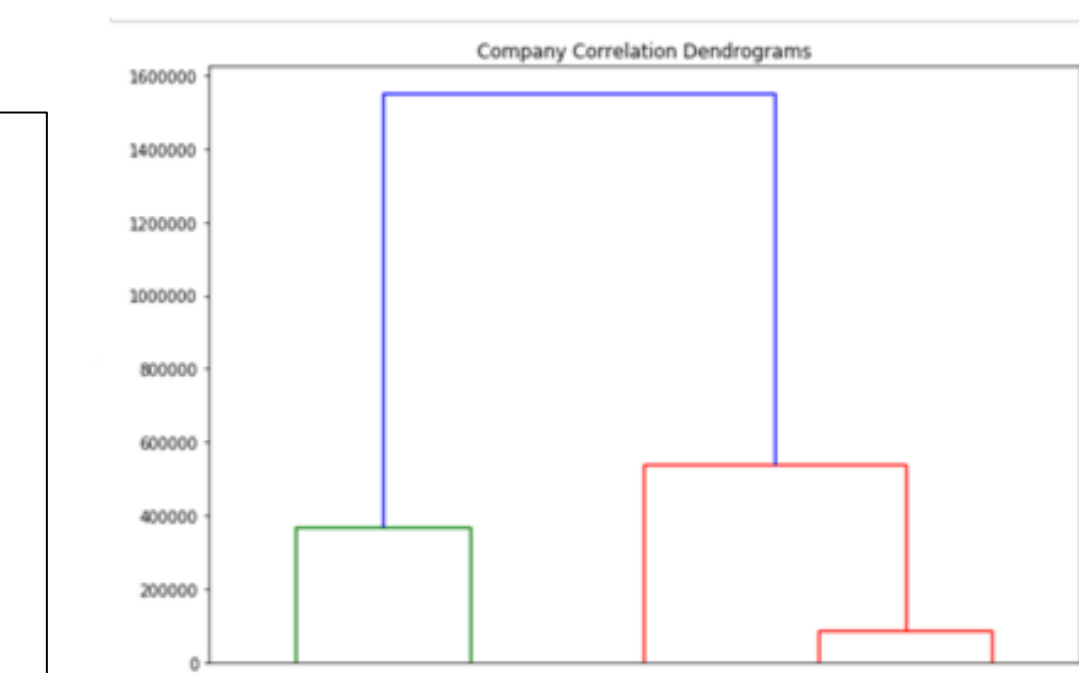- (See figure 2.4)
- (See figure 2.5)

Figure 2.4

Researched Private Equity Firms
- Found PE Firm's that generated the most revenue for UPS to understand the investment model
- Researched: Investment Criteria, Types of Portfolios, & Preferred Industries

Variable Analysis
- By analyzing the patterns in the variables, we were able to conclude which variables had the greatest influence in determining revenue for UPS
- Important Variables: Revenue, Preferred Industry, Year on Year Growth

Identification of Target Variable
- In order to create a model, we need to be able to predict a value. Using the research, we determined which value we will predict
- Target variable: UPS Net Amount Generated

Figure 2.5

WEB APP DEVELOPMENT

WHAT DOES THE WEB APP DO?
- The web app provides an interface to the users to use our predictive model

HOW WAS THE APP CREATED?
- The web app provides an interface to the users to use our predictive model

WHAT FEATURES DOES THE APP HAVE?
- Sidebar containing text and numerical input
- Displays the revenue the company will generate for UPS
- Displays the accuracy of the model and a graph showing the comparison between how close the predicted values are to the actual graph
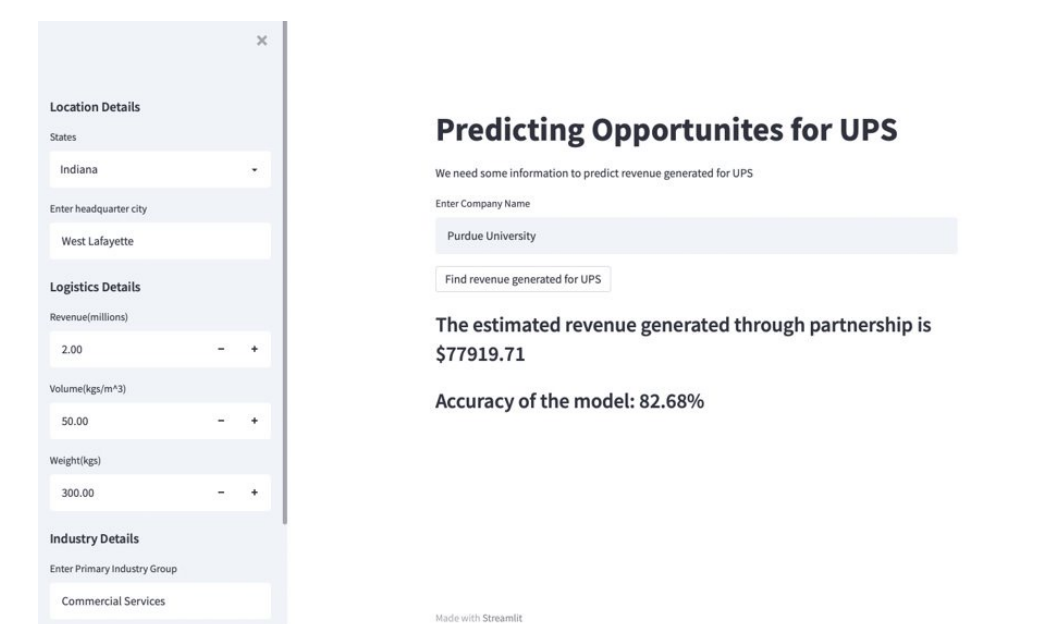
Predicting Opportunites for UPS

The estimated revenue generated through partnership is $77919.71

Accuracy of the model: 82.68%

Figure 2.3 - WebApp

CONCLUSIONS

- Started with a data set with the goal of creating a model
- Conducted market research
- Created a model that provides an 83% accuracy and can only be improved with more data being added
- Delivered a web app to UPS encoded with the model so they could input a company and get an estimate
- Determined that the model was as accurate as we could make it given the data set

FUTURE GOALS/RECOMMENDATIONS

- Consider analyzing the additional data that will be given to team, which will ultimately enhance the model
- Consider uploading data set and return information for the whole data set
- Add more columns of information to the dataset

ACKNOWLEDGEMENTS