Prediction of Small Business Health

Data Science Institute
COLUMBIA UNIVERSITY

Akanksha Rajput, Ankita Agrawal, Gurkanwar Singh, Sheetal Reddy
Chaitanya Dara,Peter Deng
Adam Kelleher

Data Science Capstone Project with Capital One

Introduction

The goal of this project is to predict the health of small businesses for forthcoming quarters using financial data obtained from their SEC filings.

In our use case, business health is quantified by the following three financial metrics :

1. Profit Margin 2. Operating Margin 3. Free Cash Flow

These standard measures of profitability and free cash available for capital expenditures are good indicators of the health of the business. This would be useful in evaluating a company's performance and making informed decisions about businesses that are safe to lend credit to or invest in.

Dataset Description

The dataset consists of 1400 small businesses that are part of Vanguard Small-Cap Index, with financial features and macroeconomic features as listed in the following table.

Features	Macroeconomic Features	Target
Price to Earnings Ratio	Unemployment Rate	Free Cash Flow (CF)
Price Cash Flow Ratio	CPI	Operating Margin (OM)
Price Book Ratio	GDP CURRY Idx	Profit Margin (PM)
Dividend Yield	Financial Stress Idx	EBITDA (dropped)
Market Valuation	Consumer Confidence Idx	
	Jobless Claims	
	ISM Manufacturing Idx	

Table 1. Dataset features and targets

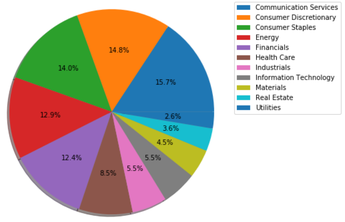

Fig 1. Industry wise composition of Vanguard Small-Cap Index Fund

Methodology

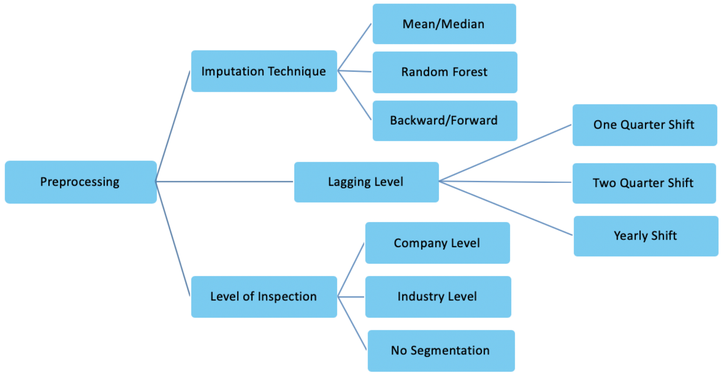

Fig 2. Experiments on pre-processing techniques

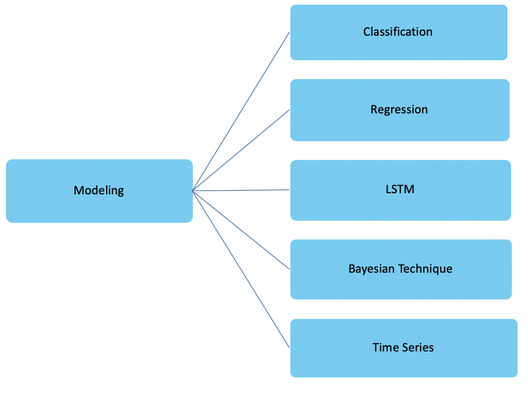

Fig 3. Experiments on modeling techniques

Results

These are the results of the best performing models under each of the aforementioned categories.

Technique	Metric	Score	Specification
Classification	Accuracy	88.34%	LDA Classifier
Regression	MAE	13.05 (with std 37.4)	Decision Tree Regressor
LSTM	Loss on Test Data	0.95	Adam Optimizer, Dropout 0.5
Bayesian	Accuracy	Didn't converge	Prior: Gaussian, Cauchy
Time Series	MSE	219842	SARIMA(1,1,1)* (1,1,0,12)

Table 2. Results of modeling techniques

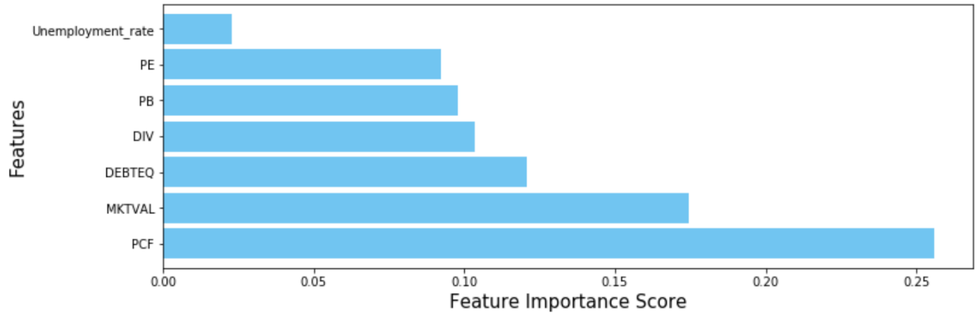

Fig 4. Top 7 Important Features

Conclusions

Our best model achieves around 90% accuracy which means this model can be used as a supporting tool for analysis by banks or individuals who are interested in lending credit to or investing in small business firms.

Acknowledgments

We would like to thank our faculty advisor Adam Kelleher and extend our gratitude to the team at Capital One (Peter Deng and Chaitanya Dara) for their continuous support and mentorship.

References

- **https://mc-stan.org/rstan/index.html**
- **https://quickbooks.intuit.com/r/financial-management/5-financial-kpis-gauge-business-health/**
- **https://www.sec.gov/edgar.shtml**