

# Forecasting Stock Returns via Supervised Learning

Samuel Ilic  
ilic@wisc.edu

Meenmo Kang  
mkang84@wisc.edu

Jonathan Santoso  
jsantoso2@wisc.edu@wisc.edu

## Abstract

We apply several machine learning models to the prediction of stock prices, specifically the price of IBM stock. Decision trees such as Random Forest, Bagging, Boosting, and Linear Regression are conducted within our research as part of a comparative analysis in an attempt to arrive at the best model. Mean square error (MSE) is used as an indicator to evaluate the accuracy of the prediction. Through this process we arrived at the best models from each method whose  $R^2$  is the highest and MSE is the lowest.

## 1. Introduction

A primary reason why people are more passionate about analysis in financial fields than in other areas is that it involves massive amounts of money. Various methodologies for predicting stock market activities have been developed over a long time, particularly via quantitative analysis by mathematicians in recent years. Moreover, as computing devices and computing ability have improved, machine learning has begun to be introduced in financial analysis. Compared to traditional statistical methods, machine learning has more flexibility to approximate random movements of data.

We therefore expected to arrive at enhanced predictions through several machine learning models, representing improvements over not only existing econometric approximations but also other market prediction methods.

The ultimate goal of our research is simply to find the best model to predict future stock returns. Decision trees have been selected for our research including Random Forest, Bagging, Boosting, and Linear Regression. These were highlighted in the class and we found these models to be a more viable and proper approach than other models for this case.

We specified a single stock; IBM was chosen for the project. IBM has a wealth of stock data history as it was long ago listed on NSYE and its price has been fluctuating, which implies there are many predictive factors, unlike the stock of consistently skyrocketing tech companies such as the FAANG (Facebook, Amazon, Apple, Netflix

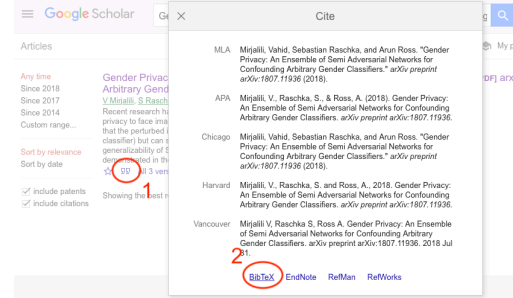


Figure 1. Example illustrating how to get BibTeX references from Google Scholar as a 1-column figure.

Method	Accuracy
Method 1	$70 \pm 3 \%$
Method 2	$76 \pm 3 \%$

Table 1. This is an example of a table.

and Google) group.

### 1.1. Subsection

You can use paragraphs or subsections to further structure your main sections. This is an example of a subsection.

**This is a paragraph title.** This is an example of a paragraph.

## 2. Related Work

Related work should be discussed here. This is an example of a citation [1]. To format the citations properly, put the corresponding references into the bibliography.bib file. You can obtain BibTeX-formatted references for the "bib" file from Google Scholar (<https://scholar.google.com>), for example, by clicking on the double-quote character under a citation and then selecting "BibTeX" as shown in Figure 1 and Figure 2.

Table 2 shows an example for formatting a table.

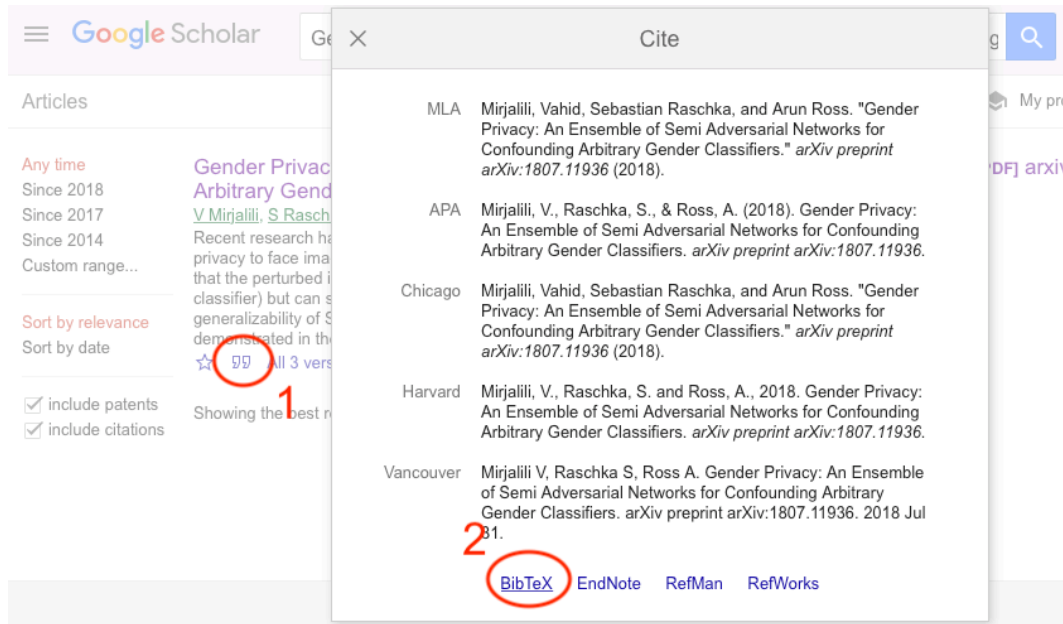


Figure 2. Example illustrating how to get BibTeX references from Google Scholar as a 2-column figure.

### 3. Proposed Method

Describe the method(s) you are proposing, developing, or using. I.e., details of the algorithms may be included here.

### 4. Experiments

Describe the experiments you performed. You may want to create separate subsections to further structure this section.

#### 4.1. Dataset

Briefly describe your dataset in a separate subsection.

#### 4.2. Software

Briefly list (and cite) software software you used.

#### 4.3. Hardware

If relevant, list hardware resources you used.

### 5. Results and Discussion

Describe the results you obtained from the experiments and interpret them. Optionally, you could split "Results and Discussion" into two separate sections.

### 6. Conclusions

Describe your conclusions here. If there are any future directions, you can describe them here, or you can create a new section for future directions.

### 7. Acknowledgements

List acknowledgements if any. For example, if someone provided you a dataset, or you used someone else's resources, this is a good place to acknowledge the help or support you received.

### 8. Contributions

Describe the contributions of each team member who worked on this project.

### References

- [1] V. Mirjalili, S. Raschka, and A. Ross. Gender privacy: An ensemble of semi adversarial networks for confounding arbitrary gender classifiers. *arXiv preprint arXiv:1807.11936*, 2018.