

Asset Management

PROJECT PLAN

Team 13

Principal Global Investors - Client
Chinmay Hedge – Faculty Advisor

Carter Scheve — *Communications Lead*
Nathan Hanson — *Project Progress Tracker/Manager*
Caleb Utesch — *Meeting Scribe*
Jack Murphy — *Research Analyst*
Samuel Howard — *Lead Engineer*
Alex Mortimer — *Project Manager*

sddec18-13@iastate.edu
<http://sddec18-13.sd.ece.iastate.edu/>

Revised: 2/11/2018, Version 1

Table of Contents

1	Introductory Material	5
1.1	Acknowledgement	5
1.2	Problem Statement	5
1.3	Operating Environment	5
1.4	Intended Users and Intended Uses	5
1.5	Assumptions and Limitations	6
1.6	Expected End Product and Other Deliverables	6
2	Proposed Approach and Statement of Work	7
2.1	Objective of the Task	7
2.2	Functional Requirements	7
2.3	Constraint Consideration	7
2.3	Previous Work And Literature	7
2.4	Proposed Design	8
2.5	Technology Considerations	8
2.6	Safety Considerations	9
2.7	Statement of Work	9
2.8	Possible Risks And Risk Management	9
2.11	Project Proposed Milestones and Evaluation Criteria	10
2.12	Project Tracking Procedures	10
2.13	Expected Results and Validation	10
2.14	Test Plan	10
3	Project Timeline, Estimated Resources, and Challenges	11
3.1	Project Timeline	11
3.1.1	First Semester	11
3.1.2	Second Semester	11
3.2	Feasibility Assessment	12
3.3	Personnel Effort Requirements	13
3.4	Other Resource Requirements	14
3.5	Financial Requirements	14

4 Closure Materials	14
4.1 Conclusion	14
4.2 References	Error! Bookmark not defined.
4.3 Appendices	Error! Bookmark not defined.

List of Figures

Figure 1 (section 3.1, pg. 11): Project Timeline

List of Tables

Table 1 - Expected Deliverables6

Table 2 - Personnel Effort Requirements13

List of Symbols

No symbols that need to be further defined were used in this document.

List of Definitions

PGI: Principal Global Investors

BK_P: Book to Price

X12M_Ret: Investment's Twelve-Month Return

1 Introductory Material

1.1 ACKNOWLEDGEMENT

Our clients at Principal have extensive experience in big data analysis and machine learning and have made it clear that they will be willing to help us in any way possible. They have already helped by providing access to learning resources such as tutorials and readings on topics like the ones mentioned above.

Our faculty advisor, Chinmay Hegde, will also be a valuable asset to developing the proposed product. As an expert in data processing and machine learning, he provides advice and guidance towards what we need to learn to deliver a successful product.

1.2 PROBLEM STATEMENT

Investment analysis at Principal Financial Group currently relies on human calculation, using a variety of models and inputs. These statistical models are proven and effective, although the dependence upon human-given inputs and calculations is both inefficient and unreliable. Various steps of the statistical analysis process can be automated, which would remove most of the potential for human error, and reduce overhead costs by making accurate statistical modeling and prediction more accessible.

Our proposed solution makes use of our extensive background in computational sciences to implement a software approach to multi-factor statistical analysis. We aim to create a system which aids in the creation and management of a profitable investment portfolio based upon well-defined statistical models and machine learning algorithms. Such a system would not only increase profits for portfolio owners, but it would also reduce risk by eliminating erroneous human action and increasing decision-making speed in a volatile stock market. The requirements given to us for the project are quite open at the moment, so our solution goals are open as well, but we have enough to start towards a functioning deliverable.

1.3 OPERATING ENVIRONMENT

The product is expected to be run on a device using Python version 3.6 or higher. Additional libraries required include Scikit-learn, Pandas, and Numpy. No environmental hazards are expected beyond those inherent in running a computer.

1.4 INTENDED USERS AND INTENDED USES

Our product has two potential user groups. The first is the investment analysts at Principal, who have little to no experience with programming techniques. The other user

group is the data analysts that are employed by Principal, who have a large amount of programming experience, along with statistical analysis and data mining.

The ideal intended use of our final product will be to provide a forecasting model that can predict with 50-60% accuracy whether or not a factor will outperform the current market. Another potential use of our final product could come even if our models are unsuccessful in predicting market behavior. Not being able to generate accurate models with the data that we were given is telling enough, and could be further analyzed to identify potential changes that would be beneficial to make in Principal's current forecasting model.

1.5 ASSUMPTIONS AND LIMITATIONS

Assumptions:

- Provided data will be of a consistent format
- Provided data will be valid and accurate
- Data will be open and accessible
- Users will have a working understanding of input and output data

Limitations:

- Tools used to research and produce model shall not exceed \$50
- The model shall be written primarily in Python
- Final predictive model must use machine learning techniques

1.6 EXPECTED END PRODUCT AND OTHER DELIVERABLES

Table 1 - Expected Deliverables

Deliverable	Description
Software Models	The main deliverable of our project is a series of software models used to predict stock market factor performance. These models include a system to input data, train the model, and display their results in a human-readable manner.
Model Analysis	The secondary deliverable is an analysis of the model's performance throughout time. We will note how different market conditions affect the accuracy of the model. Additionally, special note will be made for performance for major events, such as the 2008 stock market crash.
Documentation	The last deliverable will be the documentation of the design process used for each model. Additionally, we will include the reasoning behind each model selection and tuning process. This is to help the client investigate methods for future growth and development.

All deliverables are expected by early December.

2 Proposed Approach and Statement of Work

2.1 OBJECTIVE OF THE TASK

The objective of the proposed project of Asset Management is of two parts. One is a simulation/model to be produced. The goal here is to create a model that can (help) predict performance of certain sets of stocks. This is the backbone of the second and more indirect part of the objective, which more of a service. We will provide this model to our clients and portfolio managers will use it to help serve their clients in a more accurate way. We are providing a product and the backbone of a service for our clients to provide.

2.2 FUNCTIONAL REQUIREMENTS

- Models report processing time and accuracy
- Results are displayed in a human-readable format
- Models only use data from a certain time period to predict future behaviors
 - Example: The model does not use stock market data from 2005 to make predictions on the stock market for 2002
- Summary of models gives concrete statistics for performance of each individual model, along with a comparison of each and a recommendation for which to use in similar future tasks.

2.3 CONSTRAINT CONSIDERATION

Given there are no relevant IEE standards on machine learning, we will use the Pip8 and our client's coding standards in our design. However, in the case of a conflict, the client's standards will supersede those of Pip8. Additionally, there should be no ethical issues regarding our project.

2.3 PREVIOUS WORK AND LITERATURE

Investment analysis is a task that has been taken on by many thousands of smart minds over the course of time. Every investor that has worked in the stock market has tried to find the pattern of growth and success to gain an advantage in future deals. From what we have learned from research on the topic, other companies in the industry have software systems to help predict the behavior of stocks but, like Principal, still rely on human analysis for their larger-scale decisions.

Our product is different from projects like these in the sense that ours analyzes the success and overall performance of individual factors of investments, instead of an entire investment's overall success. In addition to the individual models, we will deliver analyses for each model, describing its effectiveness in the market along with its advantages and shortcomings.

2.4 PROPOSED DESIGN

As we discussed with our client, there are many solutions and many was to come to a solution. Thus, to cover our bases and help our clients feel better, we are going to design a few solutions that differ a little bit. Since the problem is open ended and our client is okay with us doing this, we are going to have 3-5 different analytic models. Each model will do the same thing: give predictions on the success of stocks in the near future. The differences of each model and possible solution are contained in the implementation. There are many types of models that are well known that we will use to our advantage. Each one will differ in what data we put into it, and how we tweak the internals. To put it shortly, there is really only one direction we can go to solving this and that is a constraint from our client. Our designs are not high level, but the designs are what we plan out when we are testing and analyzing data. There are almost no alternatives. Our client has given us enough creative room and time so that whatever we come up with will be a solution in some way. There is a possibility we fail in designing any possible solution. In this case, we will gain valuable insight from that and report those results.

2.5 TECHNOLOGY CONSIDERATIONS

The proposed technology to be used for our system is Python, a high-level scripting language which has been long used by statisticians for modeling and analysis. Python is a very strong choice for the system being developed; it has access to a multitude of extensive open-source mathematical modeling and statistical analysis libraries. Not only are these libraries specifically developed for purposes very similar to our own, but they are also available free of charge. Python is strong for nearly every aspect of our proposed solution. It offers utilities for easy collection and interpretation of data, as well as graphical modeling of said data. It also provides various interfaces for manipulation of data for statistical analysis purposes. It has libraries specifically for machine learning. As far as software goes, Python is probably the best tool to approach this problem.

The biggest weakness that we may encounter is a simple lack of experience with the technology. Each of us has stronger background with other software technologies. This, however, is relatively easy to overcome, as a general experience with other languages and software libraries makes it easy to learn and utilize new tools.

Alternative tools at our disposal include the R programming language, which is a language designed specifically for statistical modeling, and not much else. While we may give this some consideration, both our client and faculty advisor have suggested that we devote our focus on Python.

2.6 SAFETY CONSIDERATIONS

There are no safety concerns with the applicable product.

2.7 STATEMENT OF WORK

Research

1. Objective - The objective of this task is to gain the necessary knowledge in order to provide the optimal model for our project.
2. Approach - The task approach involves using the resources available from Principal to research machine learning and big data analysis.
3. Expected Results - The expected results from our research will be comprehension of machine learning algorithms in order to effectively use them in our models.

Model Development

1. Objective - The objective of model development is to provide a working model that will be able to effectively forecast stock market trends.
2. Approach - The approach will involve taking our information gained from research and implementing it into a working model.
3. Expected Results - A working model that can successfully forecast stock market trends based on given data would be a expected and desired result.

Testing

1. Objective - The objective of testing will be to thoroughly test and analyze our developed models to see if the desired result is achieved.
2. Approach - tuning the various parameters in our models to analyze the results.
3. Expected Results - Many of the initial results will probably not be what we need them to be at, but further testing will hopefully provide information to feasibly predict trends in the market.

2.8 POSSIBLE RISKS AND RISK MANAGEMENT

One of the main risks simply involves a lack of knowledge for the area of machine learning. Many of us have no previous experience working with machine learning. This project will be a good chance for us to gain knowledge and experience in this area, but initially it will be difficult to adjust with the amount of learning required. There is also the risk that the models we develop will not provide sufficient information for Principal to properly predict future outcomes from our research.

As far as risk management techniques, we are looking to implement a risk management log that will enable us to identify and track possible risks for this project.

2.11 PROJECT PROPOSED MILESTONES AND EVALUATION CRITERIA

The first key milestone in our proposed project will most likely be if and when we begin to produce models that are useful in identifying patterns in the data we were given.

Throughout our project implementation we will be making many different statistical models based on the data we've been given. Once we begin to notice patterns or trends within the different parameters as they relate to the given factors, our strategy will most likely begin to change significantly based on these identified patterns.

Testing any patterns we identify will involve modifying parameter values that are used as input for our model. We will need to check with multiple different input values for any parameters involved that the output that is being produced is at least close to what we would expect it to be. If multiple input values don't give outputs close to what we expect them to be, then this most likely an indication that the pattern we thought we identified might not be accurate.

2.12 PROJECT TRACKING PROCEDURES

Our team uses the Gitlab project tracking software to monitor project progress. This is a great choice, as the system incorporates work tasks, TODO items, and progress reports in a clear, concise, effective format that bolsters our work environment.

2.13 EXPECTED RESULTS AND VALIDATION

The desired outcome of our project is to deliver a software tool that analyzes performance reports on past investments and generates relatively accurate predictions for investments moving forward. This will be used to monitor the pool of investments on the market and inform the user about ones that could soon become profitable or dangerous to own.

To confirm that our solution works at a satisfactory level in the industry, we will run several tests with a multitude of data sets containing information on investment performance from the past two decades. After running the data through our program, we will compare the output to the actual results for the investments and find the level of accuracy attained. If our models perform better than the norm for human analytics, we will have created a successful product.

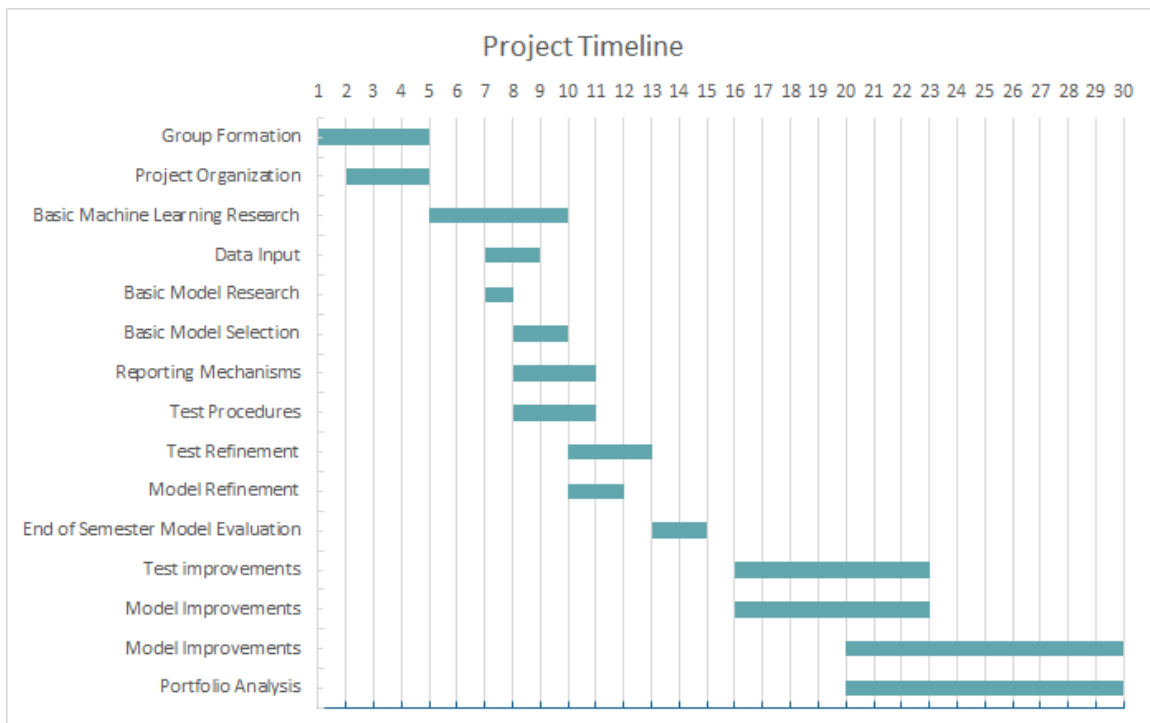
2.14 TEST PLAN

Our plan for testing is fairly straightforward. We will run a large amount of simulated investment environments with our software, monitoring the success of each at several different time periods. With all of that data, we can then compare each of the models and modeling techniques to establish which works better for this type of analysis. This will be valuable information for our client, and will allow them to carry on the work after our project terminates.

3 Project Timeline, Estimated Resources, and Challenges

3.1 PROJECT TIMELINE

Figure 1 - Timeline of Tasks



3.1.1 First Semester

The first semester will be composed of background research, exploratory data analysis, and preliminary model selection. We will be poking around the data, trying to see which predictors work best, and try and reason out why for each. Datasets with many different variables, like stock market data, is incredibly hard to visualize, so we will find the best through trial and error.

3.1.2 Second Semester

The second semester will be taken up with taking the best model(s) found previously, and refining it. We will try to make the models 'better' according to whatever standards we decide upon when refining our tests. Near the end of the semester we will do the portfolio analysis, which entails seeing how good our model would do throughout the time period presented in the dataset. Special note will be taken for periods of volatility, such as the 2008 stock market crash. The analysis, model, and documentation will be presented to the client at the end of this semester.

3.2 FEASIBILITY ASSESSMENT

This project is somewhat of a challenge to assess; however, we see the plan as quite feasible. We want the project to succeed to the fullest. The success of this project would equate to this state: a set of models that have a relatively high accuracy in forecasting the market for a certain set of stocks. These models will be available to people within PGI to use at their disposal. There is a substantial amount of challenges, but we see all of the challenges enumerated and have planned for them to the best of our ability. The first challenge is the barrier to entry for this area of expertise. Most of the group currently has little to no experience in machine learning and data analysis. We also have some barrier to entry with the tools that we'll be using to develop and test our models. Another challenge will be in the development of our models. The big question concerns how accurate we can get with our models and how influential they will be on the current strategies of investment analysis. The challenge is that there are many ways in which a model can be influenced. Which factors or qualities are the most influential are the ones we have to find, which is a challenge in itself. A lot of parts of the development of this project initially seem to be challenges, but minor ones at best. The major ones are listed here and we expect to overcome them at each point in time that they come up.

3.3 PERSONNEL EFFORT REQUIREMENTS

Table 2 - Personnel Effort Requirements

Planning and Documentation	Developing a project plan and design document, as well as other documents such as a risk management log and project timeline, will enable us to stay on track throughout the duration of this project. Our goal is to make these documents detailed and clear in order to limit the amount of roadblocks going forward.
Research	Quite a bit of our effort will be spent in research. In order to develop a successful model, a significant amount of research will need to be done. Research will be focused on machine learning and exploratory data analysis.
Issue Tracking	Gitlab will be used for issue tracking of our project. Proper use of Gitlab will be imperative for our success in this project. Each team member will be able to post issues as well as set the necessary date for completion.
Model Development	Since developing several viable and accurate models is the overall end goal of our project, model development will most likely be where the bulk of our software development efforts go towards. This includes testing out different machine learning algorithms with our proposed models, as well as the work associated with creating the models themselves.
Testing	Once models are developed, extensive testing will need to be done. This will most likely be handled by each group member as there will likely be more than one model to test.

3.4 OTHER RESOURCE REQUIREMENTS

This project is purely software related, which implies that we will not need any hardware resources. We have confirmed this to be true. Aside from resources we already have, namely computers, there are no resources that we need to complete the project.

3.5 FINANCIAL REQUIREMENTS

The nature of the project makes it so that there are very little, if any, financial dependencies. For each part: learning, developing, testing, and deploying is predicted to cost no money at. However, we do allow a little bit of money as a just in case. There are some places for learning and testing that might cost money, but other than that, we see no cost incurrence in the future for this project.

4 Closure Materials

4.1 CONCLUSION

The stock market is largely regarded as unpredictable, yet companies like Principal strive to achieve consistent profits for clients by attempting to forecast stock market trends by analyzing patterns in different factors of stocks. Factor analysis at principal is largely done by humans, which can be inaccurate, biased, and above all, inefficient. We hope to develop a useful resource that will be widely utilized at principal. In our work, we expect to provide insight on the viability of various statistical models based on factor analysis through software tools. Software provides consistent and reliable outputs for a set of inputs, and has already been used in high-volume stock trading and many other areas in the financial industry. By introducing our tool to Principal's business model, we have to potential to greatly add to the company value.