Faculty of Engineering & Information Technology

32513 - Advanced Data Analytics Algorithms

“INVEST - ASSIST “ Investment Prediction Sytem

Pitch and Plan for Solving a Domain Problem

Subject Coordinator: Dr. Jun Li

By

Vishnu Mohan Edala Dhanraj 13069909

Table of Contents

Video Pitch : https://www.youtube.com/watch?v=YSDckq-MFC0&rel=0

GITHUB : https://github.com/VishnuMohanED/UTS_ML_2019_ID13069909/blob/master/ML_Ass3_13069909.pdf

INVEST- ASSIST

INTRODUCTION

The main purpose of the project is to make an app and web-based application tool for customers mainly to achieve their financial goals and needs. This document will provide a detailed report on the proposal. The further sections will show case the aim and objectives of the project, significance and innovation involved, information about the model developed, timeline, budget and personnel required to complete the project, challenges and steps taken to complete them.

AIM

The main aim of this project is to develop 'INVEST - ASSIST' an Investment Prediction System which provides all the personal assistance a user needs while making better investment options and also to attain their financial goals and needs.

OBJECTIVES

The main objectives of the Invest - assist project is,

- To analyze different types of investments like Growth investments, shares, property, bond, Gold and defensive investment types and extract their data to develop decision support models.
- Build an app repository for all the above-mentioned investment categories and provide a single stop solution.
- To develop a predictive data model and provide recommendations to the users based on their preferences.
- provide complete analysis to the customer in both short term and long-term investments.

Further this document helps in providing timeline, budget and personnel required to pull of this project without any discrepancies which will help to cater the potential investors.

BACKGROUND

Investment is nothing but commodities or general good purchased today with a plan of using them in the future or at the time of crisis. In the similar way financial investment refers to putting aside a fixed amount of money or an asset that will grow into a large sum or provides gain over a stipulated time frame. But the way of investment and planning involved is different between the users. There involves various type of methods or types to invest, some of them are real estate (both residential and commercial property), equities, bond, stock Mutual funds and Gold/silver ornaments. The current trend shows that customers are shuffling between different investment consultants and different set of applications for each type of investment to have an idea about the trend in the market and make decisions. It's hard for any customer to handle many applications and draw insights out of it.

Each investment companies uses different independent application tools based on their specialization, Eg. For Stocks the plus500 application gives a complete overview and

assistance, for mutual funds- FINRA fund analyzer provides great tool, for Gold and ornaments - various different assistance providers, for trading X-Trade Brokers (XTB) for UK and several for other countries etc. All these above stated applications or tools are mostly confined to a certain category and are mostly independent to other. Therefore, there is not any correlation prevailed between any of these investment tools. This provides the way of paying independently for each consultation practices which is expensive and troublesome to customers.

Therefore, the main objective of this project is to provide a centralized repository for all the investment options together and provide assistance for making better investment options. For instance, the user will have his profile with all his background information and financial status details. Considering the forecast made by this information with respect to equity or bond or investment thorough various tools, it's hard to compare and comprehend. Hence the investment options correlated together can make the user a single stop solution to compare the forecast their revenue and risk in each of the investment options. Further, the user can keep an eye on all the investments made , current market trend, the relationships made between multiple investments and their respective current asset value.

The main step of making this application is to collect heaps and heaps of data of different investment options made, their respective predictive models and decision-making algorithms involved in those investment options. This project requires intense research on all the investment options, comparisons made between them and detailed analysis of their characteristics and relations between them. Subject matter experts should be on board and their assistance in each step of making the application work is recommended in building the predictive models.

Predictive analysis is dominant and got support of worldwide organizations and has global market value projected to about 10.95$ billion by 2022. Since the model that we propose for Invest- Assist mainly provides suggestions to the customer to make their financial decisions, predictive analysis plays an important role in building this model, as it deals with wide range of data technologies including bigdata, data mining, statistical and machine learning models. Thus, by measure of all these, the application provides forecast and trends in current situation and predictions that may occur at specific time based on the parameters applied. Each investment category follows independent prediction models, for instance (Osman Hegazi, Omar S. Soliman and Mustafa Abdul Salam, 2013) provides algorithm that integrates particle swarm optimization (PSO) and least square support vector machine (LS-SVM) for predicting the stock prices. Thus, decision making model varies based on the investment types. In this project we will be using multiple prediction models for making the entire system.

RESEARCH PROJECT

SIGNIFICANCE

The project proposed possess lot of significance and out of all , the important aspects to look out are listed as follows,

- Central repository for all investment options.
- Ease of use for customers

- Cost Effectiveness
- Provides assistance based on user needs and financial status.

Central repository for all investment options

Earlier, users and consultants have to rely on independent tools for each category of investment options, get license to use them. But Invest- Assist will provide assistance to all types of investment options and predict their forecast.

Ease of use for customers:

This Tool will help users to look for all investment options ranging from stock, equity, bond and defensive investment options, instead of each tool for different categories. Thus providing ease to customers to analyze the trends and predict forecasts on a single tool.

Cost Effectiveness:

Invest -Assist provides assistance to make financial calls of all the investment options, thereby reducing the cost to have application each for individual needs. Cost effectiveness is the major significance of this tool, as it reduces the cost burden of users.

Provides assistance based on user needs and financial status:

This tool will have a set of preference methods running for the users while logging in for the first time. Based on his likes, preferences and current financial status. the tool will provide suggestions for all types of investments under one single roof.

INNOVATION

The idea of proposing a single investment assistance tool for all kinds of investments and financial needs is the most innovative thing in this data driven world. The proper combinations of using machine learning methods and AI concepts involved in predicting the models will provide high range of accuracy and predictions which will help the users to provide a dignified and easy means of providing financial assistance based on their preferences and type. Moreover, the system has mechanisms to analyze live data to and predict the current trend. Therefore, this system will take inputs from old and new data to predict the future trend. This will greatly attack customers to try them and in turn help the stakeholders to gain valuable input and alter based on the customers need and preferences.

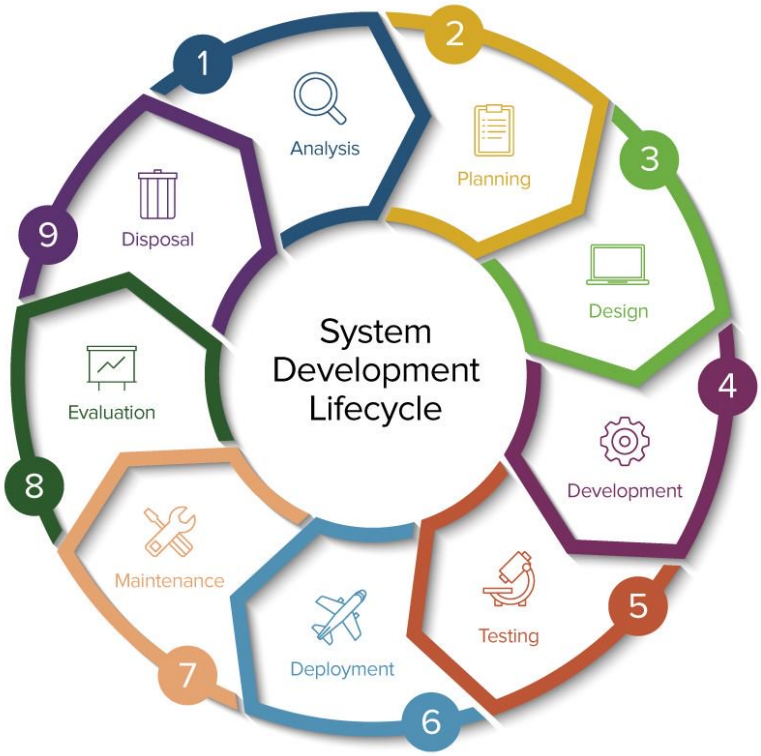
Thus, when this project goes live on floors, it will contribute and makes a great impact in today's Data driven world and help financial consultants and investors to make right and perfect decisions.

PROJECT PLAN

The Project here follows software development lifecycle and is being classified into 4 stages of 9phases in total. The four stages of the development are listed as below,

- PLAN - Careful planning and comprehensive analysis of requirements.

- DO - Build predictive models and recommendation system
- CHECK - Trial versions on floors for testing purposes
- ACT - Look for Customer feedbacks and keep on working on them for betterment and perfection.



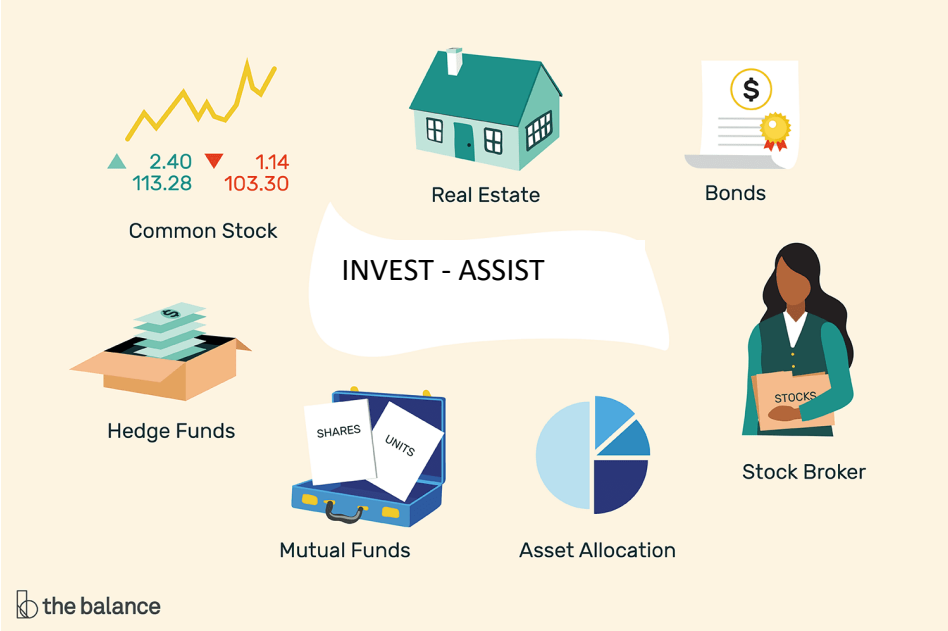
Phase 1 : PLAN

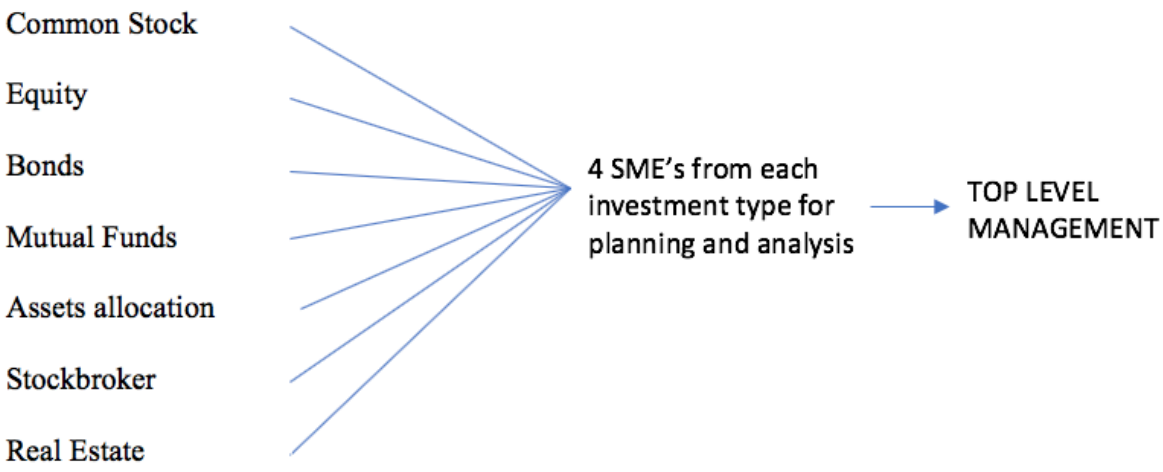
The prime and the most important step involved in this process is **Analysis** of existing technologies and their methods. The preliminary stage involves careful and complete analysis of all the tools and investment applications with their important features and target audience. This analysis will give an improved idea of what is expected for customers from each investment type or services and prepare a detailed report on them.

For instance, plus500 provides complete assistance for stocks, bitcoin transactions, trade and licensing. FINRA fund analyzer helps in providing assistance in mutual funds. The analysis involved in the first stage will give a detailed report on existing structures and likelihood of the target audience. These are considered as inputs for developing the project.

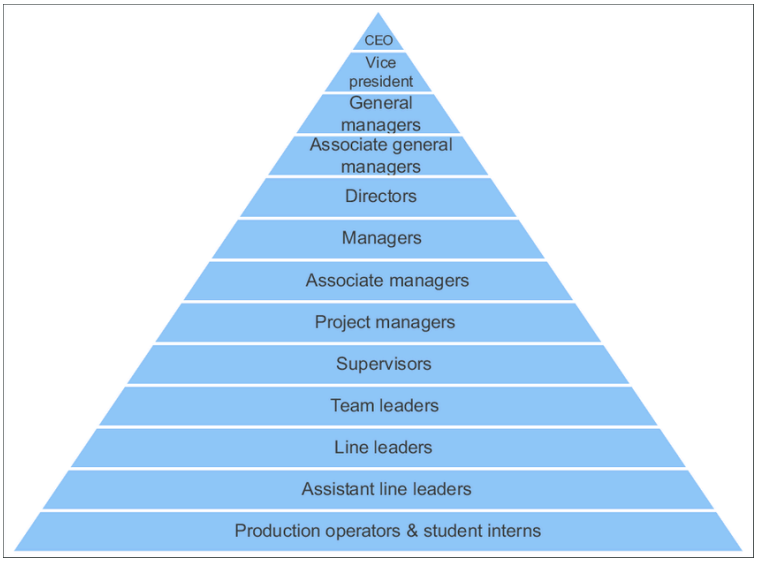
The Next step involved in **Planning**, where experts of all the financial investment types ranging from stocks, mutual funds to Gold and defensive types should be taken on board. The experts will plan and do make a complete report on the requirements and needs for commencing the project. The Experts will be giving a brief about the project's requirement, budget, personnel needed to finish the project to the project managers. A complete hierarchical structure is developed to showcase the different command centers and their leaders during the process of making the application. Each investment category will have 4 SME's working towards their domain area of knowledge, out of which 2 will provide detailed explanations on the existing standards and 2 of the senior SME's will provide enhancements or innovative concepts to the existing ones. Thus when everything is documented the SME's of all the investment areas will

submit their work to the Top-level management. Once its completed and approved, they are made to be communicated to the IT teams to commence the Core part.





Thus, the top-level management has a complete set of management team involved in for project execution and maintenance which can be shown as,



This stage is expected to be happening for two months, and the IT team will be working on the software development prerequisites which is needed to carry out the project. The Core technical team will parallelly be working on collecting the data set of various stocks and investment types details for analysis and predicting cycles. With all the preliminary stages are done and necessary information or dataset is collected the project then move onto the design and development phase.

BUILD PREDICTIVE MODELS AND RECOMMENDATION SYSTEM

DATA SOURCE

The next stage of the process involves data extraction and processing. Since this project deals with 7 investment types with massive volume of data being extracted from various sources ranging from

- Data in reserve,
- Twitter data - using official API by extracting tweets for sentimental analysis,
- ASX (Australian Securities Exchange) which provides stock details, market quotes , shares and other information with respect to investments, information
- Details of assets
- Real estate datasets. Etc.

Out of these listed above, there are some data which are dynamic in nature. For Eg. Data about stock exchange rates are not reliable as they keep on changing every few minutes of time. Certain data being extracted from Investopedia and twitter extracts are not in standard format. Technical teams should be in place to convert them into usable format. Later all these data should be confined into a single repository (Data Integration) which can be later used for analysis purposes.

ANALYSIS

Once the Data is preprocessed, integrated and all set for the next phase which is data analysis. The IT core team will start working on data analysis part where the team will start building model to predict investments. There are certain investment types like equity and bond which are quite common these days. But the real question arises when Prediction of trend of these investment types are really working? Does a common user can use it? The answer is NO. This is mainly because of the lack of user's interest and his target. Previous models dint actually look for user's personal interest and information. To overcome this, recommendations should be provided based on certain information from the user by their respective inputs gathered while accessing the application which can actually relate to their predictive models.

Moreover, both short term and long-term investments benefits suggested to the customer, based on their financial status and needs. Thus, The IT teams will be developed certain models which can suit the user needs and type of investment that the user want to make in.

DATA MANAGEMENT

Data Is the prime most thing to be kept confidential in this data driven world. Whichever the industry you work in or whatever the role that you are in, the Data is actually changing each one lives by one way or the other. Thus, the data provides wonderful outcomes and insights when handled and incase of being lost at any point it would incur losses directly. This the project deals with the confidential and sensitive information and therefore proper assurance to be made to users. To facilitate this best assuring and secured governance of data should be in place.

Separate teams to handle different set of data, for eg. Legal advisor for handling legal concerns, business development and handling team to completely handle the data and IT team to manage and provide enhancements constantly should be in place.

All these processes are touted to process for 6 months and therefore each module has to be processed, discussed and made changes at the right point as incase of any discrepancies it will be sorted whenever needed to avoid further loss in time and money.

BUILDING THE SOFTWARE

Once the Model is developed, tested and validated for predicting accuracy levels and forecast, the next phase involves developing software or application systems. We will be developing both webs based and mobile app-based interfaces which will be available both on android and IOS.

The software will be developed and tested for all kind of redundancies. Then it will be processed for various security and testing practices before it actually reaches the final phase of availability to users. To test the developed software model and get customer feedback the app version of the software is released for use. The business team will be working on checking the customer satisfactory levels, take account of both constructive and destructive feedbacks.

Then the Web based platform is made available to public, with all the necessary information about the software in the supporting documents. The IT team will coherently work on integrating the web platform and the mobile application platforms to be in sync.

When the software tool is made available to public the IT team will work on integrating it with the Cloud platforms as since this involves large amount of data and lots of space is needed to handle the software system.

DATA SOURCE

The source of data to be fed to the models are extracted from various sources. Some of them are listed below,

- ASX - Australian securities exchange which provides current stock quotes , market data, share prices to predict the trend.
- TWITTER - Extract tweets using official API for sentimental analysis and predict the trend.
- Investment and Assets - Information extracted from banks and real estate data for generating models or predicting customers preference over real-estate to cash or gold /silver investments.
- Trusted Articles - financial and stock exchange information
- Investopedia - An American website mainly focuses in financial assistance and investment suggestions.

Tools for Data Analysis and Software System

- Rapid Miner - for data preprocessing and web crawling.
- Python - Data Analysis and Visualization.
- Node.js + npm software - to design the front end of the application. The npm software has an online data base registry of public and paid for private packages. This is further being used to develop software interface systems.

Database and Management System

Database

- Apache Kafka: a distributed streaming platform developed by LinkedIn . Kafka is mainly used for building real-time data with high throughput, low latency.

Database Management System

- Spark - It uses micro-batching for real time streaming. Used to analyse large amount of data.
- Neo4j - graph database open source tool to extract the data insights and relationship between them.

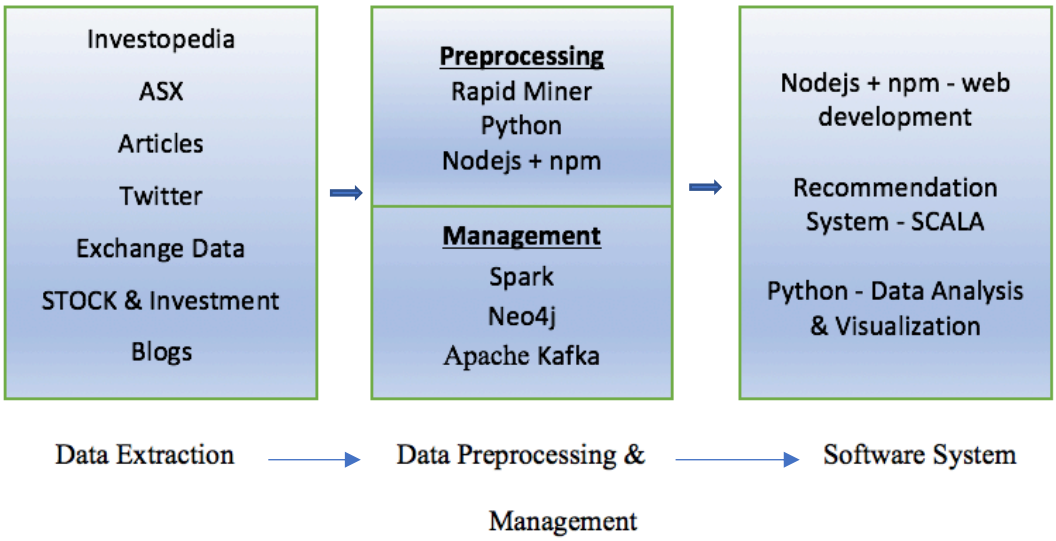
Programming Language

- SCALA - Highly preferred along with apache spark to build recommendation systems.

Outcomes

- App Based software named Invest-Assist, which is released first on mobile platforms for trails to have customers thoughts. The application will be made available on both android and iOS platforms.
- A web-based platform to access the software tool Invest-Assist which will be launched with all its necessary support documents.

The work flow the Software development is shown below,



TIMELINE

The timeline to this project is expected to run for a year, with different timelines for completing each module and deliver the final product. The time taken to complete each module is shown from the following table,

Phase	Time Frame (In Months)
Preliminary Works in Understanding business needs	2 months
Data Extraction	Both these phases start together
Data preprocessing & integration	1 Month
Building Classifier Models	2 Months
Testing Accuracy of Predictions	1 Month
Building Software Model	3 Months
Pilot Version Deployment	1 Month
Launching full version and making Review Report	2 Months

The detailed time frame for each tasks and subtasks involved in this project is shown in the appendix section.

EXPECTED OUTCOMES AND BENEFITS

The main target audience to this project is financial Investors and consultants. This application will benefit the customers by saving their valuable time and money instead of relying on multiple tools. This tool will act as a consolidated platform to access multiple investment types for better forecasting model and cater their needs.

Based on the user's requirement with Ownership -Eg. Stocks, real estate properties, savings, and private equity, Lending - Bonds, TIPS and Deposits, Alternatives - Such as silver and gold deposits, this tool will provide outcomes as follows,

- Recommendations based on their type of investment, estimated profit along with short term and long-term benefits.
- Sentimental analysis using twitter feed to find the current trend.
- Provides estimated income trend.
- Constant advice notifications from financial experts.
- Manage assets over a wide range of finance types.
- Automated notification on investment options.
- Expense report and Tax file assistance.
- Keep a complete record on all your investments and records made.

BUDGET

This Segment will briefly discuss about the Budget involved in making this project. This is further classified into three categories based on the level of usage and need,

- Personnel
- Infrastructure
- Office Space

The total budget touted is 2,465,940 A$ which is schedule for a period of 365 days.

Personnel			
		A$	
Designation	**Approx. Working Days**	**Charge Per Day**	**Total Cost**
Project Head(1)	250	1500	375000
Project Supervisor (5)	250	1500	375000
Subject Matter Experts (14) (Contract)	75	2000	150000
Subject Matter Experts (14)(fulltime)	150	2000	300000
Data Analysts(25)	70	2500	175000
web Developer(10)	50	1500	75000
Technical Support Team(10)	30	700	21000
Marketing Team(5)	30	500	15000
		Total :	1486000

Infrastructure			
		A$	
Product Type	**Approx. Count**	**Cost per Item**	**Total Cost**
Laptops	200	2000	400000
Servers & Database	5000	50	250000
Software Licenses	10	500	5000
Screens	50	250	12500
Other Storage Devices	25	100	2500
		Total :	670000

Office Space			
		A$	
Category	**No of Days**	**Cost per Day**	**Total Cost**
Rent	365	400	146000
Bills	365	50	18250
		Total :	164250

Total Budget in A$	
Total Cost (Infrastructure + Office + personnel)	2320250
Contingencies (10%)	145690
Final Budget (Total Cost + Contingencies)	2465940

PERSONNEL

Subject Matter Experts (SME's)

The ultimate success of this project mainly depends on the kind of technical team being involved in making this tool. A slight deviation from the expected work can cost lot of money and time involved. The managers and business analysts will be involved right from the day the project commences. Short meetings and quick meetings should happen constantly to ensure that all the technical teams are working together, and everyone know about what actually the process involved in it. Considering the subject matter experts, 2 of them will be leaving when the data extraction and analysis is done. Where the other 2 will be with us till the end of the

project. These 2 SME's will keep an eye on the project status and assist us in case of any deviations.

IT - Technical & Support Team

All the IT personnel will be working till the end of the project. Where a few of them will be retained even after the project output is made on floors. This is to ensure the application is working as expected and incase of any deviations the technical team will provide all the assistance needed to make this project tool back on track. A technical support team will be laid with all the knowledge about the project being shared to him. This team will be constantly be working on providing enhancements and support in case of any issues.

Business & Marketing Team

The Business and marketing team will be working throughout the project. Right from the planning stage till the project launch, all the steps are considered with a business point of view. Proper marketing strategies will be used to reach out to the target audience and make this project a huge success.

REFERENCES

Osman H., Omar S., Mustafa Abdul S., 2013, A Machine Learning Model for Stock Market Prediction, *International Journal of Computer Science and Telecommunications,* vol. 4, viewed 7 october 2019

https://www.researchgate.net/profile/Mustafa_Abdul_Salam/publication/259240183_A_Machine_Learning_Model_for_Stock_Market_Prediction/links/00b7d52a93b06f2b35000000/A-Machine-Learning-Model-for-Stock-Market-Prediction.pdf

Savia L., 2018, *Building Recommendation System with Scala and Apache Spark [Tutorial],* Packt, viewed 8 october 2019, https://hub.packtpub.com/building-recommendation-system-with-scala-and-apache-spark-tutorial/

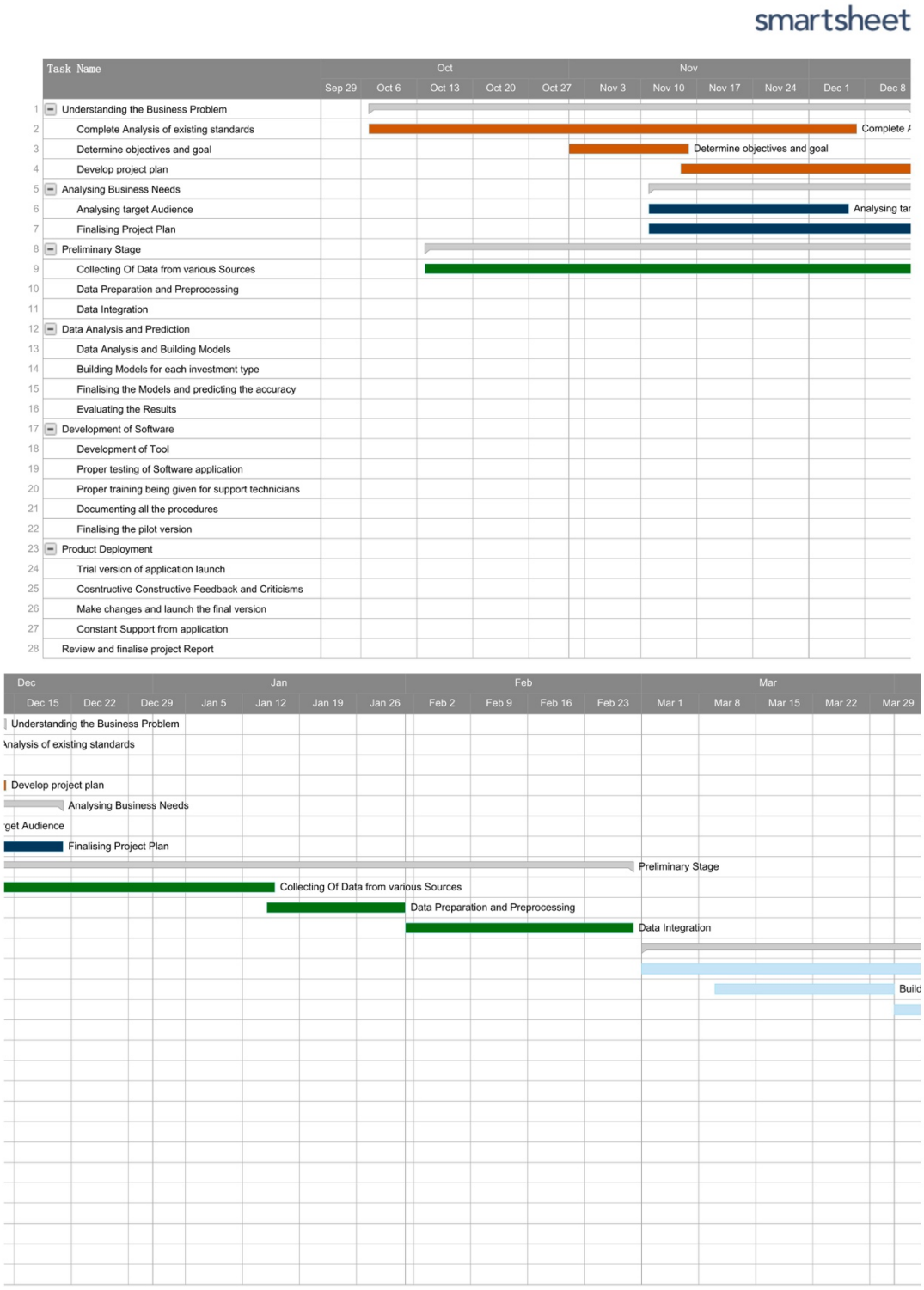
Christian P., *What is Apache Kafka? Why is it so popular? Should you use it?,* TechBacon, viewed 8 october 2019, https://techbeacon.com/app-dev-testing/what-apache-kafka-why-it-so-popular-should-you-use-it

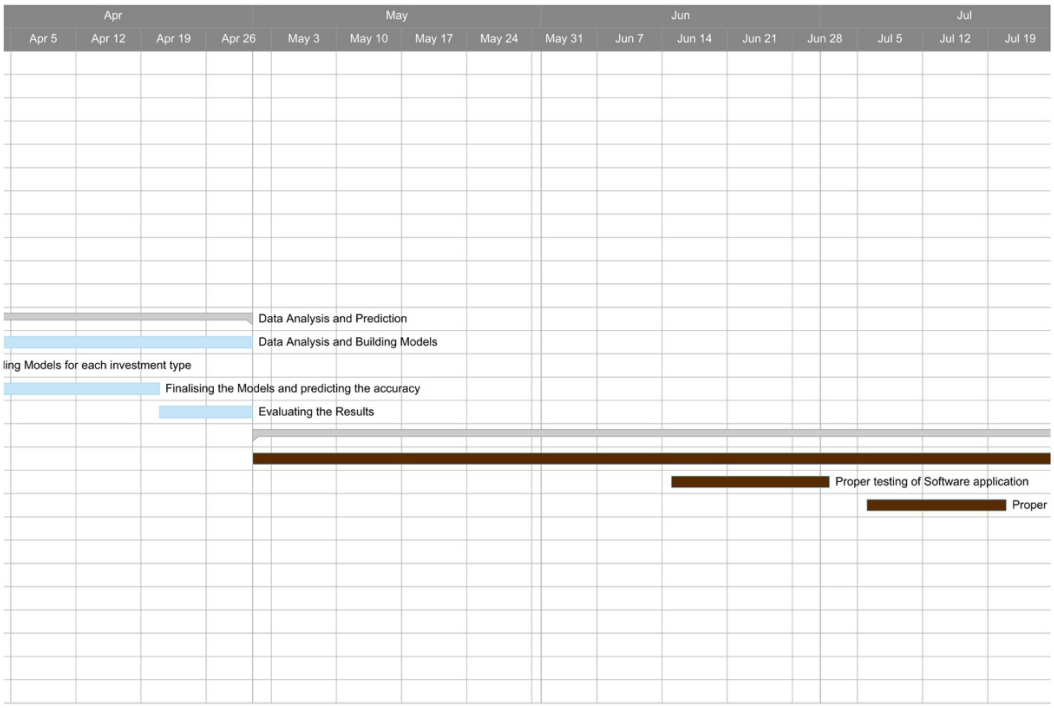
Joseph T L., 2006, Investment Analysis for Real Estate Decisions, *Journal of Property Investment & Finance,* vol. 24, no. 3, pp. 268-269

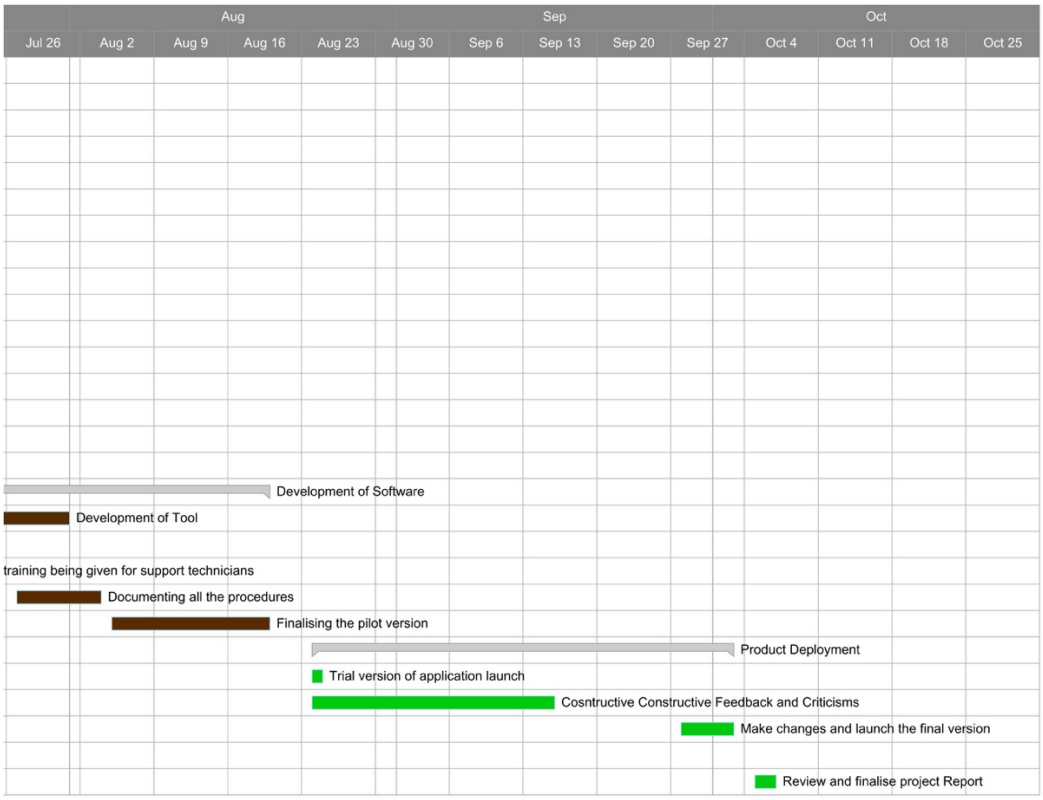
APPENDIX

GANTT CHART

The project is planned for a year and the respective processes and timeline for each tasks are shown here using the Gantt chart.




Apr
May
Jun
Jul
Apr 5
Apr 12
Apr 19
Apr 26
May 3
May 10
May 17
May 24
May 31
Jun 7
Jun 14
Jun 21
Jun 28
Jul 5
Jul 12
Jul 19
Data Analysis and Prediction
Data Analysis and Building Models
ling Models for each investment type
Finalising the Models and predicting the accuracy
Evaluating the Results
Proper testing of Software application
Proper


Aug
Sep
Oct
Jul 26
Aug 2
Aug 9
Aug 16
Aug 23
Aug 30
Sep 6
Sep 13
Sep 20
Sep 27
Oct 4
Oct 11
Oct 18
Oct 25
Development of Software
Development of Tool
training being given for support technicians
Documenting all the procedures
Finalising the pilot version
Product Deployment
Trial version of application launch
Cosntructive Constructive Feedback and Criticisms
Make changes and launch the final version
Review and finalise project Report