

A Project Phase I Report On

RECOMMENDATION FOR

E – COMMERCE USING MACHINE LEARNING ALGORITHMS

Submitted in Partial fulfillment of the requirement for the award of the
degree of

BACHELOR OF TECHNOLOGY IN

COMPUTER SCIENCE AND ENGINEERING

SUBMITTED

By

19671A0597

SUNDARAM

19671A0591

REBBAVARAPU VINITHASREE

19671A0559

BOBBILI SHYAM

19671A05A0

PRATEEK BHURE

Under the esteemed guidance of

Mr. HIMAGIRI

ASSISTANT PROFESSOR



Department of Computer Science and Engineering

Accredited by NBA

J.B. Institute of Engineering & Technology

(UGC AUTONOMOUS)

(Affiliated to Jawaharlal Nehru Technological University, Hyderabad)

Bhaskar Nagar, Yenkapally, Moinabad Mandal, R.R. District, Telangana (India)-

500075

2022-2023

J.B. INSTITUTE OF ENGINEERING & TECHNOLOGY

UGC AUTONOMOUS

(Accredited by NAAC & NBA, Approved by AICTE & Permanently affiliated by JNTUH)

Yenkapally, Moinabad Mandal, R.R. Dist-75 (TG)

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING



CERTIFICATE

This is to certify that the project report entitled “**RECOMMENDATION FOR E – COMMERCE USING MACHINE LEARNING ALGORITHMS**” submitted to the Department of Computer Science and Engineering, J.B Institute of Engineering & Technology, in accordance with Jawaharlal Nehru Technological University regulations as partial fulfilment required for successful completion of Bachelor of Technology is a record of Bonafide work carried out during the academic year 2022-23 by,

SUNDARAM **19671A0597**

R.VINITHASREE **19671A0591**

B.SHYAM **19671A0559**

PRATEEK BHURE **19671A05A0**

Internal Guide

Mr.HIMAGIRI

ASSISTANT PROFESSOR

Head of the Department

Mr. G. SREENIVASULU

ASSOCIATE PROFESSOR

External Examiner

J.B. INSTITUTE OF ENGINEERING & TECHNOLOGY
UGC AUTONOMOUS
(Accredited by NAAC & NBA, Approved by AICTE & Permanently affiliated by JNTUH)
Yenkapally, Moinabad Mandal, R.R. Dist-75 (TG)

(Accredited by NAAC & NBA, Approved by AICTE & Permanently affiliated by JNTUH)
Yenkapally, Moinabad Mandal, R.R. Dist-75 (TG)

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING



DECLARATION

I hereby certify that the Major Project report entitled “**RECOMMENDATION FOR E – COMMERCE USING MACHINE LEARNING ALGORITHMS**” carried out under the guidance of our faculty, **Mr.HIMAGIRI Assistant Professor** is submitted in partial fulfillment of the requirements for the award of the degree of **Bachelor of Technology in Computer Science and Engineering**. This is a record of Bonafide work carried out by me and the results embodied in this major project report have not been reproduced or copied from any source. The results embodied in this project report have not been submitted to any other university or institute for the award of any other degree or diploma.

Date: / /

SUNDARAM **19671A0597**

R. VINITHASREE 19671A0591

B. SHYAM **19671A0559**

PRATEEK BHURE 19671A05A0

ACKNOWLEDGEMENT

At outset I express my gratitude to almighty lord for showering his grace and blessings upon me to complete this Mini Project. Although my name appears on the cover of this book, many people have contributed in some form or the other to this project development. I could not have done this Project without the assistance or support of each of the following.

First of all, I am highly indebted to **Dr. P. C. KRISHNAMACHARY**, Principal for giving us the permission to carry out this Major Project.

I would like to thank **Mr. G. SREENIVASULU**, Associate Professor & Head of the Department of COMPUTER SCIENCE AND ENGINEERING, for being moral support throughout the period of the study in the Department.

I am grateful to **Mr.HIMAGIRI**, Assistant Professor COMPUTER SCIENCE AND ENGINEERING, for his valuable suggestions and guidance given by him during the execution of this Project work.

I would like to thank Teaching and Non-Teaching Staff of Department of Computer Science and Engineering for sharing their knowledge with us.

SUNDARAM	19671A0597
R. VINITHASREE	19671A0591
B. SHYAM	19671A0559
PRATEEK BHURE	19671A05A0

CONTENTS

<u>SL. NO</u>	<u>NAME OF THE TOPIC</u>	<u>PAGE NO</u>
	ABSTRACT	
1.	INTRODUCTION	1
2.	LITERATURE SURVEY	2
3.	SYSTEM ANALYSIS	
	3.1 Existing System	4
	3.2 Drawbacks of Existing System	4
	3.3 Proposed System	4
	3.4 Advantages of Proposed System	5
	3.5 Software Development Life Cycle Model (SDLC)	5
	3.6 Project Implementation Plan	6
4.	SOFTWARE REQUIREMENT SPECIFICATIONS	
	4.1 Functional Requirements	7
	4.2 Non-Functional Requirements	7
	4.3 Software Requirement Specifications	7
	4.4 Hardware Requirement Specifications	8
5.	SYSTEM DESIGN	
	5.1 System Architecture	9
	5.2 Design Tool Used	10
	5.3 UML Diagrams	
	5.3.1 Use Case Diagram	20
	5.3.2 Class Diagram	21
	5.3.3 Sequence Diagram	22
	5.3.4 Activity Diagram	22
6.	BIBLIOGRAPHY	23

LIST OF FIGURES

Sl.No	Desperation of Figure	Page No
1	SDLC	5
2	System Architecture	9
3	Support vector machine	10
4	SVM Scenario-1	11
5	SVM Scenario-2	11
6	SVM Scenario-3	12
7	SVM Scenario-4	13
8	SVM Scenario-5	14
9	Random Forest	16
10	Use Case Diagram	20
11	Class Diagram	21
12	Sequence Diagram	22
13	Activity Diagram	22

ABSTRACT

Due to the huge expand in global markets and financial transactions , the importance of E-commerce has grown significantly , so to achieve fully functioning , scalable , reliable ,efficient, secure , and user-friendly E-commerce system , Adequate system analysis and design procedures are essential. On the other hand The internet vast growing has totally altered the way most companies work. The Internet has created a means for e-Commerce the internet has created a way to provide a unique avenue for companies and customers to sell and buy goods and services. When creating an ecommerce website, there are several goals that must be taken into account, one of them is how to Increase the site's efficiency to ensure customer turnout and thus achieve the required material profits. There are several methods that are followed to increase the efficiency of the site. One of these methods is the recommender system. This paper produces a study overview of the value of recommended systems and thoroughly analyzed Collaborative Recommender System (CF) techniques with its techniques., which presents proposals to customers according to their interests, making it easier for the customer to search and thus choose the goods that suit him.

1. INTRODUCTION

E-commerce is defined by Whinston and Kalakota as “the process of purchasing and marketing of information, goods and services by computer networks; which are mainly being the internet[1]. However, the term is used by others to include not only the previously mentioned buying and marketing, on the other hand the internet technologies usage, such as email and intranets, to share or exchange information either within the enterprise or with external stakeholders. [2]. Other meanings agree that e-commerce often denotes the acts, techniques and principles are needed to promote the integration of electronic communication into the business climate. A more inclusive e-commerce’s concept would be: Capacity to achieve commercial contacts and full agreements Exchange of goods and services between the concerned parties through the use of instruments, regulations, methods and techniques for electronic, internet and/or telecommunications. More businesses are producing online presence to keep track of their competitive advantage due to the massive growth of the internet network. For any organization to succeed domestically or internationally, the internet has become a critical instrument. The Internet's exponential growth has completely altered the workings of most companies. There are various types of products in an inclusive online shopping site like Alibaba, and Amazon and so on , therefore there is a problem for client to find out a suitable item from all the others, which in turn would influence the clients' interest for purchasing which in turn reduce the sales of trades. Therefore, an appropriate recommender system will be very essential for the clients and businesses of an inclusive e-commerce web site. In order to increase the performance of ecommerce system , recommendation system is used which is depends in most existing systems only on purchasing information .A recommender system acquires information from a client and recommends goods that it will find most valued from among the existing products. Recommendation Systems are considered as software tools and techniques to suggest products for customers via taking into account their favorites in an automated way. The provided suggestions aimed to provide customers in numerous decision-making ability. Recommendation systems are embedded in various fields such as knowledge recovery (IR), machine learning, decision support systems (DSS), and text classification. Through recommending consumers with potentially important or useful items, to deal with the problem of information overload (IO), these systems are used. They have proven to be useful IO processing tools for online clients and have become one of the most common and powerful e-commerce tools.

2. LITERATURE SURVEY

H. song study examines whether macro-level opportunity indicators affect cyber-theft victimization. Based on the arguments from criminal opportunity theory, exposure to risk is measured by state-level patterns of internet access (where users access the internet). Other structural characteristics of states were measured to determine if variation in social structure impacted cyber-victimization across states. The current study found that structural conditions such as unemployment and non-urban population are associated with where users access the internet. Also, this study found that the proportion of users who access the internet only at home was positively associated with state-level counts of cyber-theft victimization. The theoretical implicate of these findings are discussed.

P.Alaei, proposed model, called Network Anomaly Detection using Active Learning (NADAL) involves an offline and an online step. The selected dataset is preprocessed in an offline fashion. The NSL-KDD dataset contains instances labeled with the attack types. During the preprocessing step, the attacks are divided into four categories: DoS, Probe, R2L, and U2R. Furthermore, there are four classifiers at the respective layers of attacks. Thus, the preprocessing carried out using Weka selects the appropriate features for each classifier.

The selected features are then given to the feature filtering module in NADAL.

M. Saber, described the results achieved using the JAM distributed data mining system for the real-world problem of fraud detection in financial information systems. For this domain clear evidence that state-of-the-art commercial fraud detection systems can be substantially improved in stopping losses due to fraud by combining multiple models of fraudulent transactions shared among banks. Demonstrated that the traditional statistical metrics used to train and evaluate the performance of learning systems, (i.e. statistical accuracy or ROC analysis) are misleading and perhaps inappropriate for this application. Cost-based metrics are more relevant in certain domains, and defining such metrics poses significant and interesting research questions both in evaluating systems and alternative models, and in formalizing the problems to which one may wish to apply data mining technologies.

M.Tavallee, focused on recent research around anomaly detection, specifically, the work published during the period of 2000–2008. Analyzed three major components in each study that are critical for the evaluation and comparison of the intrusion-detection techniques. These components include the employed datasets, the characteristics of the performed experiments and the methods used for performance evaluation.

A.S. Ashoor, have designed a network intrusion detection system based on the artificial neural networks using

Multi-Layer Perceptron (MLP) and Modified signature Apriory algorithm and the testing results of the prototype system proved the validity of the method and the advantages over other methods suggested. In the present study a more general problem is considered in which the attack type is also detected such as smurf, teardrop,etc. This feature enables the system to suggest proper actions against possible attacks.

M.Zamani, proposed an application of machine learning to support the processing of large datasets holds promise in many industries, including financial services. However, practical issues for the full adoption of machine learning remain with the focus being on understanding and being able to explain the decisions and predictions made by complex models. Explored ability methods in the domain of real-time fraud detection by investigating the selection of appropriate background datasets and runtime tradeoffs on both supervised and unsupervised models.

M.Chakraborty, this proposed system aims to provide a systematic and comprehensive overview of these issues and challenges that obstruct the performance of FDSs. We have selected five electronic commerce systems; which are credit card, telecommunication, healthcare insurance, automobile insurance and online auction. The prevalent fraud types in those E-commerce systems are introduced closely. Further, state-of-the-art FDSs approaches in selected E-commerce systems are systematically introduced. Then a brief discussion on potential research trends and conclusions are presented.

3.SYSTEM ANALYSIS

3.1 EXISTING SYSTEM

The packaging company that provided the data for this research has a long history of sales expertise. This expertise is captured predominantly in the intuition of sales representatives. Intuition is not easy to record and disseminate across an entire sales force, however, and thus one of the company's most valuable resources is inaccessible to the broader organization. As a result, the company tasked this team with extracting the most important factors.

3.2 DRAWBACKS OF EXISTING SYSTEM

- Most prior work in this space has been performed by private companies.
- The research in the field is typically unavailable to the public.
- The academic work that does exist either is related to forecasting aggregate sales instead of scoring opportunity level propensity, or is based on custom algorithms that fall outside the standard tools used by data scientists in industry.

3.3 PROPOSED SYSTEM

To clean the data and cut out inessential information prior to modeling, the team first filtered out all entries when the system was formally launched for the company. Variables with a high percentage of null values were then excluded to ensure a sufficient sample size. The remaining variables were further screened based on potential importance determined by conversations between the team and key company stakeholders. Additionally, data exploration resulted in several opportunities for feature engineering and custom variables to capture potential influence not captured in the default fields.

3.4 ADVANTAGES OF PROPOSED SYSTEM

- Count of the number of fields completed in one record.
- Count of the number of tasks for the customer account associated with an opportunity.

3.5 SOFTWARE DEVELOPMENT LIFE CYCLE MODEL (SDLC)

The spiral model is an SDLC model that combines elements of an iterative software development model with a waterfall model. It is advisable to use this model for expensive, large and complex projects. In its diagrammatic representation, we have a coil having many cycles or loops. The number of cycles varies for each project and is usually specified by the project manager. Each spiral cycle is a stage in the software development process.

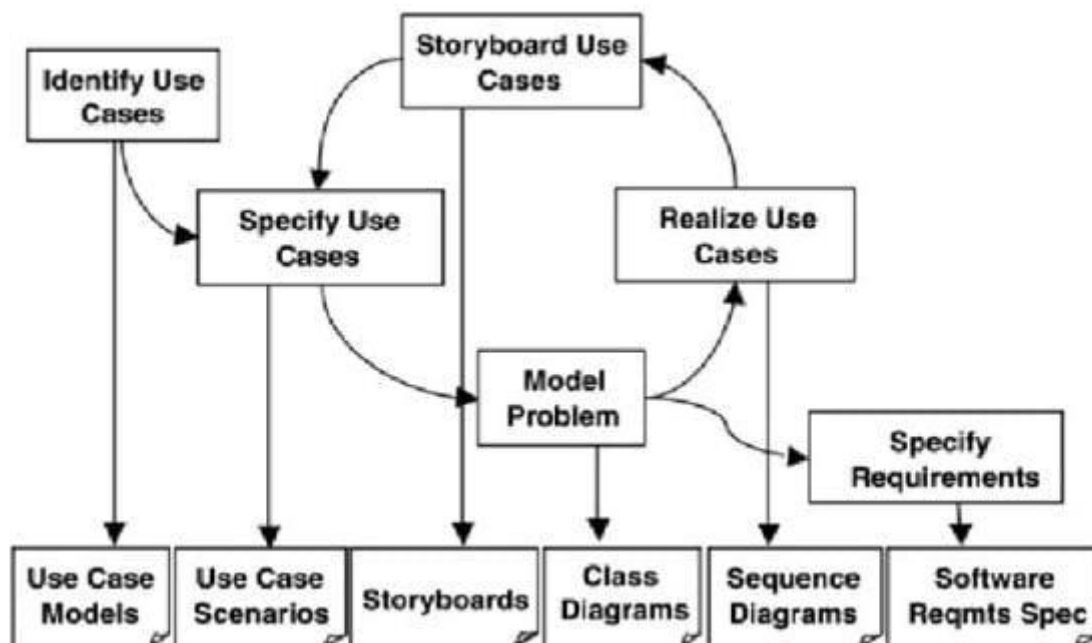


Fig 3.5 – SDLC

3.6 PROJECT IMPLEMENTATION PLAN

To better achieve the objective of predicting open opportunities, it would be prudent to capture and model how opportunity fields change over time, perhaps via periodic snapshots. This way, the company would be able to make predictions at different stages in the opportunity lifecycle.

Another important application of these kinds of prediction models is to assist in determining where to invest sales time and resources for business planning optimization.

Predictions from accurate models are also worth rolling up into aggregate sales forecasts and adjusting existing “bottom-up” methods. Before these applications would be addressed however, data ops resources would be required to perform a number of critical tasks: continue building and tuning the model for better accuracy, establish a cadence around maintaining the models and incorporating new kinds of information, and connecting with the other business units to understand strategic priorities for operationalization.

4.SOFTWARE REQUIREMENT SPECIFICATIONS

The project involved analyzing the design of few applications so as to make the application more users friendly. To do so, it was really important to keep the navigations from one screen to the other well ordered and at the same time reducing the amount of typing the user needs to do. In order to make the application more accessible, the browser version had to be chosen so that it is compatible with most of the Browsers.

4.1 FUNCTIONAL REQUIREMENTS

- Graphical User interface with the User.

4.2 NON-FUNCTIONAL REQUIREMENTS

For developing the application the following are the Hardware Requirements:

- Processor: Pentium IV or higher
- RAM: 4 GB
- Space on Hard Disk: minimum 1 tb

4.3 SOFTWARE REQUIREMENT SPECIFICATIONS

- **Operating System:** Windows
- **Coding Language:** Python, Django
- **Script:** Anaconda
- **Database:** NSL-KDD

4.4 HARDWARE REQUIREMENT SPECIFICATIONS

- **Processor:** Pentium –III
- **Speed:** 2.4 GHz
- **RAM:** 512 MB (min)

- **Hard Disk:** 20 GB
- **Floppy Drive:** 1.44 MB
- **Monitor:** 15 VGA Color

5. SYSTEM DESIGN

5.1 SYSTEM ARCHITECTURE

Collaborative filtering filters information by using the interactions and data collected by the system from other users. It's based on the idea that people who agreed in their evaluation of certain items are likely to agree again in the future.

The concept is simple: when we want to find a new movie to watch we'll often ask our friends for recommendations. Naturally, we have greater trust in the recommendations from friends who share tastes similar to our own.

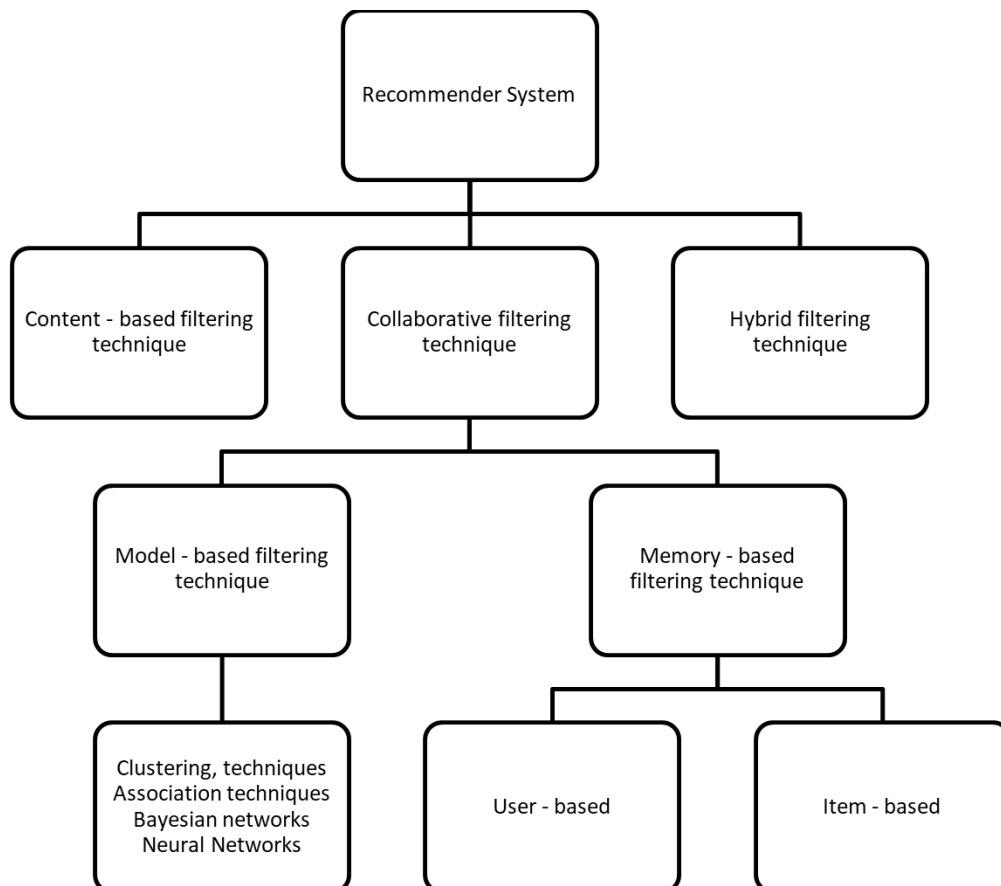


Fig 5.1 – System Architecture

5.2 DESIGN TOOL USED

In fraud app detection we have used:

- Machine learning techniques (SVM, Random Forest).
- Python for coding Python libraries (TensorFlow, keras, Scikit-Learn, Flask)
- Anaconda software to deploy.

5.2.1 MACHINE LEARNING

5.2.1.1 Support Vector Machine:

Support Vector Machine (SVM) is a supervised machine learning algorithm which can be used for both classification and regression challenges. However, it is mostly used in classification problems. In the SVM algorithm, we plot each data item as a point in n-dimensional space (where n is number of features you have) with the value of each feature being the value of a particular coordinate. Then, we perform classification by finding the hyper-plane that differentiates the two classes very well (look at the below snapshot).

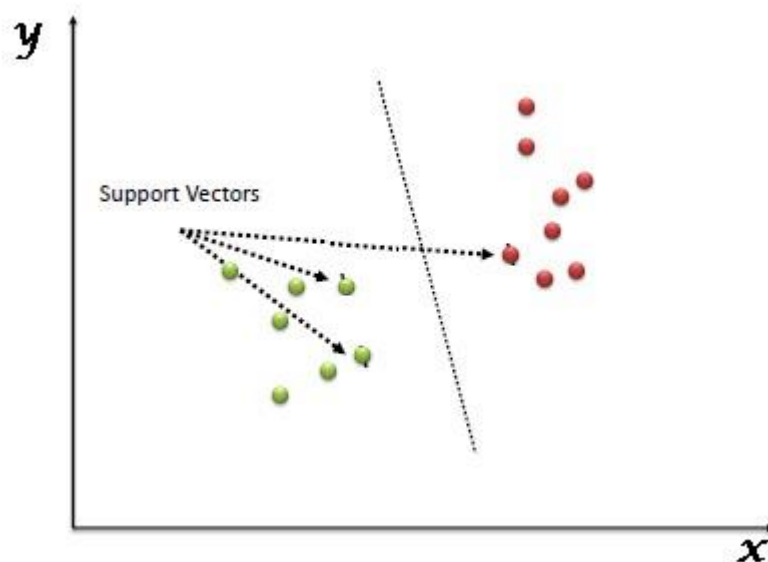


Fig 5.2.1.1(a) – Support vector machine

Support Vectors are simply the co-ordinates of individual observation. The SVM classifier is a frontier which best segregates the two classes (hyper-plane/ line)

- **Identify the right hyper-plane (Scenario-1):** Here, we have three hyper-planes (A, B, and C). Now, identify the right hyper-plane to classify stars and circles.

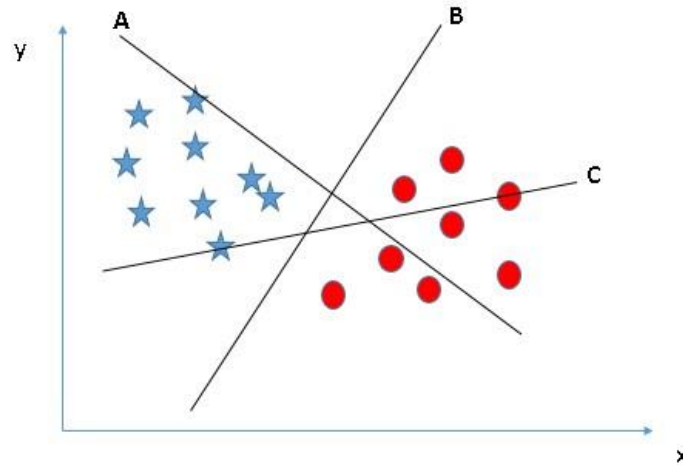


Fig 5.2.1.1(b) – SVM Scenario-1

A thumb rule to identify the right hyper-plane: “Select the hyper-plane which segregates the two classes better”. In this scenario, hyper-plane “B” has excellently performed this job.

- **Identify the right hyper-plane (Scenario-2):** Here, we have three hyper-planes (A, B, and C) and all are segregating the classes well. Now, How can we identify the right hyper-plane?

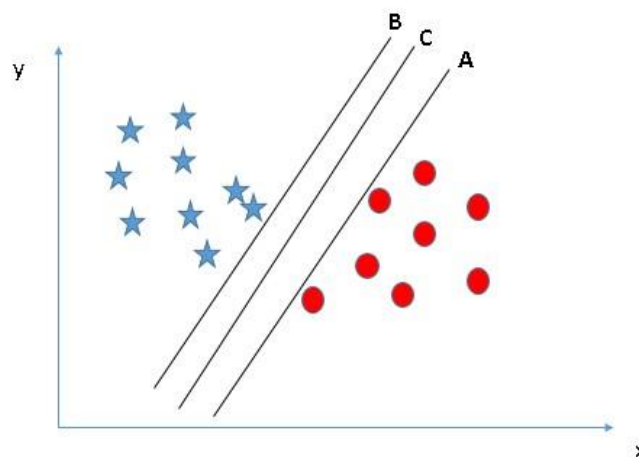


Fig 5.2.1.1(c) – SVM Scenario-2

Here, maximizing the distances between nearest data point (either class) or hyper-plane will help us to decide the right hyper-plane. This distance is called **Margin**. Above, you can see that the

margin for hyper-plane C is high as compared to both A and B. Hence, we name the right hyper-plane as C. Another lightning reason for selecting the hyper-plane with higher margin is robustness. If we select a hyper-plane having low margin, then there is high chance of miss-classification.

- **Identify the right hyper-plane (Scenario-3):**Hint: Use the rules as discussed in previous section to identify the right hyper-plane

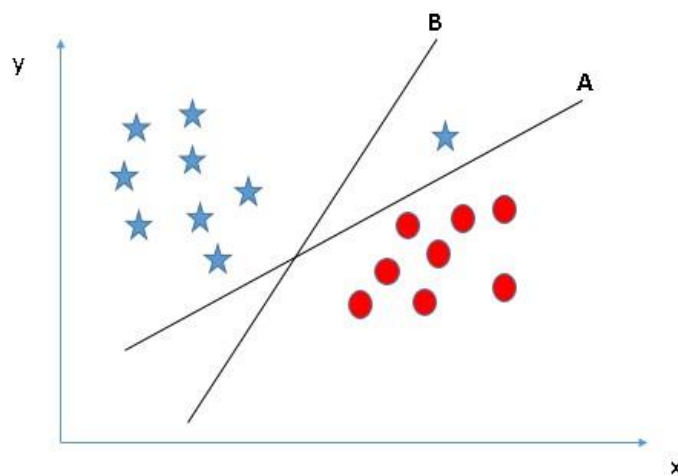


Fig 5.2.1.1(d) – SVM Scenario-3

Some of you may have selected the hyper-plane **B** as it has higher margin compared to **A**. But, here is the catch, SVM selects the hyper-plane which classifies the classes accurately prior to maximizing margin. Here, hyper-plane B has a classification error and A has classified all correctly. Therefore, the right hyper-plane is **A**.

- **Can we classify two classes (Scenario-4) ?** : Below, I am unable to segregate the two classes using a straight line, as one of the stars lies in the territory of other(circle) class as an outlier. As we have already mentioned, one star at other end is like an outlier for star class. The SVM algorithm has a feature to ignore outliers and find the hyper-plane that has the maximum margin. Hence, we can say, SVM classification is robust to outliers.

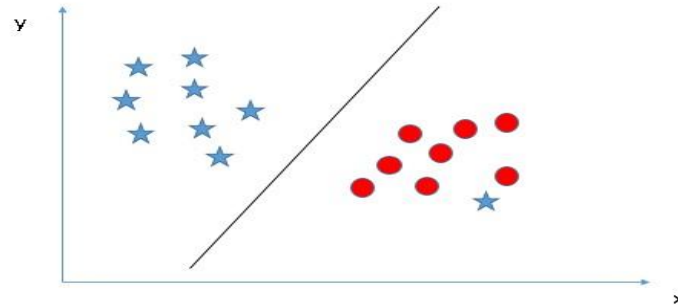


Fig 5.2.1.1(e) – SVM Scenario-4

- **Find the hyper-plane to segregate to classes (Scenario-5):** In the scenario below, we can't have linear hyper-plane between the two classes, so how does SVM classify these two classes? Till now, we have only looked at the linear hyper-plane.

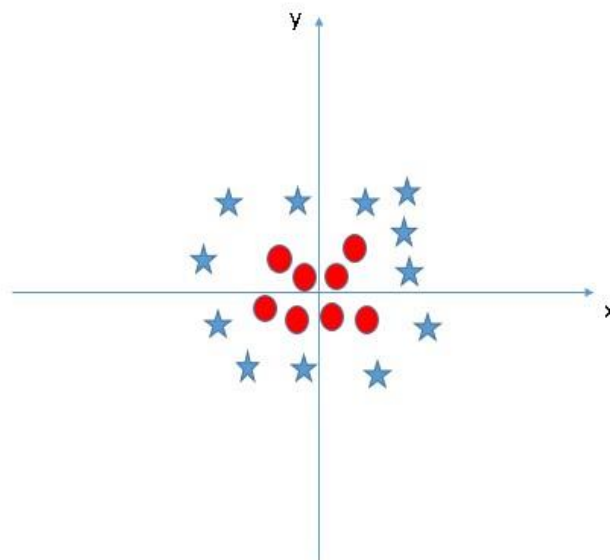


Fig 5.2.1.1(f) – SVM Scenario-5

SVM can solve this problem. Easily! It solves this problem by introducing additional feature. Here, we will add a new feature $z = x^2 + y^2$. Now, let's plot the data points on axis x and z:

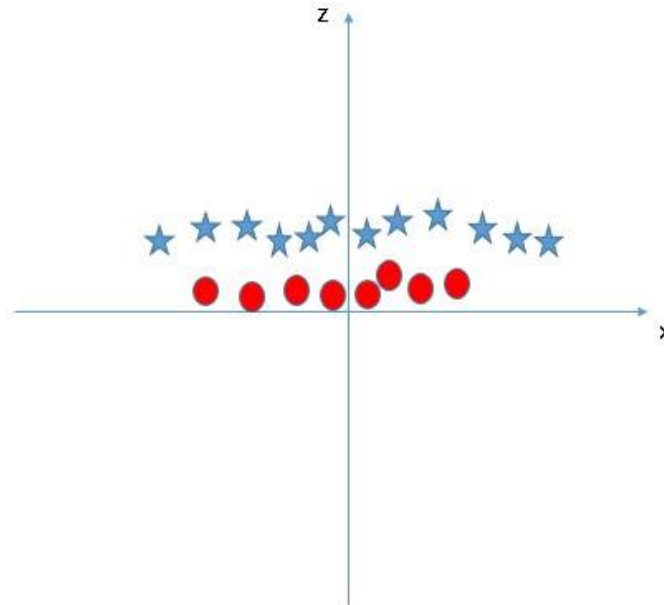


Fig 5.2.1.1(g) – SVM Scenario-5

In above plot, points to consider are: All values for z would be positive always because z is the squared sum of both x and y . In the original plot, red circles appear close to the origin of x and y axes, leading to lower value of z and star relatively away from the origin result to higher value of z . In the SVM classifier, it is easy to have a linear hyper-plane between these two classes. But another burning question which arises is, should we add this feature manually to have a hyper-plane. No, the SVM algorithm has a technique called the kernel **trick**. The SVM kernel is a function that takes low dimensional input space and transforms it to a higher dimensional space i.e., it converts a not separable problem to a separable problem. It is mostly useful in nonlinear separation problems. Simply put, it does some extremely complex data transformations, then finds out the process to separate the data based on the labels or outputs you've defined.

When we look at the hyper-plane in original input space it looks like a circle:

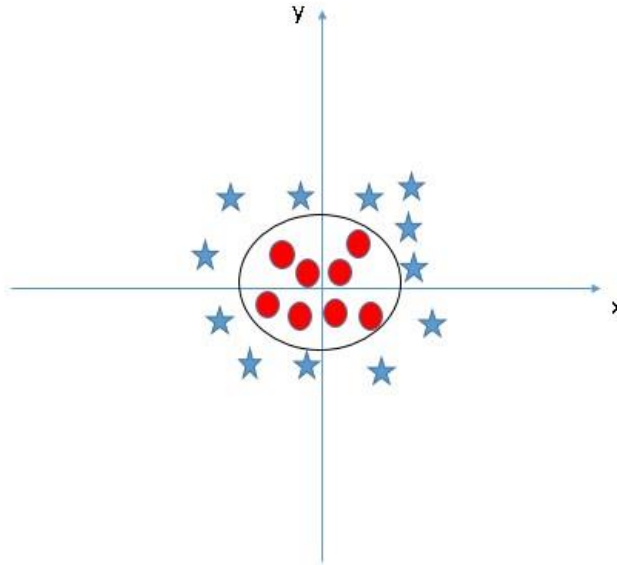


Fig 5.2.1.1(h) – SVM Scenario-5

5.2.1.2 Random Forest Algorithm

Random Forest is a popular machine learning algorithm that belongs to the supervised learning technique. It can be used for both Classification and Regression problems in ML. It is based on the concept of ensemble learning, which is a process of combining multiple classifiers to solve a complex problem and to improve the performance of the model. As the name suggests, "Random Forest is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset." Instead of relying on one decision tree, the random forest takes the prediction from each tree and based on the majority votes of predictions, and it predicts the final output. The greater number of trees in the forest leads to higher accuracy and prevents the problem of overfitting.

The below diagram explains the working of the Random Forest algorithm:

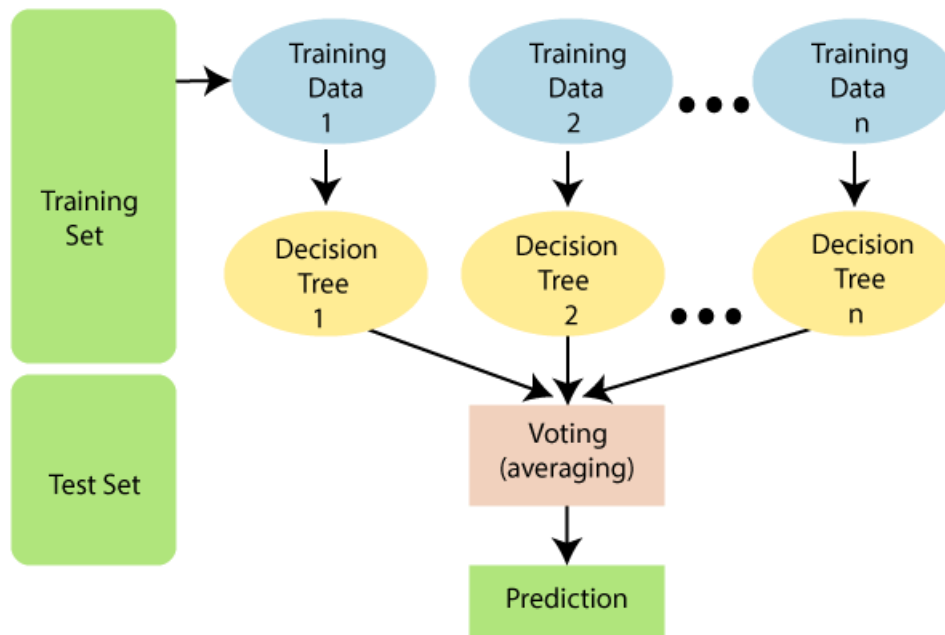


Fig 5.2.1.2– Random Forest

How does Random Forest algorithm work?

Random Forest works in two-phases: first is to create the random forest by combining N decision tree, and second is to make predictions for each tree created in the first phase.

The Working process can be explained in the below steps and diagram:

Step-1: Select random K data points from the training set.

Step-2: Build the decision trees associated with the selected data points (Subsets).

Step-3: Choose the number N for decision trees that you want to build.

Step-4: Repeat Step 1 & 2.

Step-5: For new data points, find the predictions of each decision tree, and assign the new data points to the category that wins the majority votes.

The working of the algorithm can be better understood by the below example:

Applications of Random Forest

There are mainly four sectors where Random Forest mostly used:

1. **Banking:** Banking sector mostly uses this algorithm for the identification of loan risk.
2. **Medicine:** With the help of this algorithm, disease trends and risks of the disease can be identified.
3. **Land Use:** We can identify the areas of similar land use by this algorithm.

4. **Marketing:** Marketing trends can be identified using this algorithm.

Advantages of Random Forest

- Random Forest can perform both Classification and Regression tasks.
- It is capable of handling large datasets with high dimensionality.
- It enhances the accuracy of the model and prevents the overfitting issue.

Disadvantages of Random Forest

- Although random forest can be used for both classification and regression tasks, it is not more suitable for Regression tasks.

5.2.2 PYTHON

Python is currently the most widely used multi-purpose, high-level programming language. Python allows programming in Object-Oriented and Procedural paradigms. Python programs generally, are smaller than other programming languages like Java. Programmers must type relatively less and indentation requirement of the language, makes them readable all the time. Python language is being used by almost all tech-giant companies like – Google, Amazon, Facebook, Instagram, Dropbox, Uber... etc.

The biggest strength of Python is huge collection of standard libraries which can be used for the following –

- Machine Learning
- GUI Applications (like Kivy, Tkinter, PyQt etc.)
- Web frameworks like Django (used by YouTube, Instagram, Dropbox)
- Image processing (like OpenCV, Pillow)
- Web scraping (like Scrapy, BeautifulSoup, Selenium)
- Test frameworks
- Multimedia

Advantages of Python:

1. **Extensive Libraries:** Python downloads with an extensive library and it contain code for various purposes like regular expressions, documentation-generation, unit-testing, web browsers, threading, databases, CGI, email, image manipulation, and more. So, we don't have to write the complete code for that manually.

2. **Extensible:** As we have seen earlier, Python can be extended to other languages. You can write some of your code in languages like C++ or C. This comes in handy, especially in projects.
3. **Embeddable:** Complimentary to extensibility, Python is embeddable as well. You can put your Python code in your source code of a different language, like C++. This lets us add scripting capabilities to our code in the other language.
4. **Improved Productivity:** The language's simplicity and extensive libraries render programmers more productive than languages like Java and C++ do. Also, the fact that you need to write less and get more things done.
5. **IOT Opportunities:** Since Python forms the basis of new platforms like Raspberry Pi, it finds the future bright for the Internet of Things. This is a way to connect the language with the real world.
6. **Simple and Easy:** When working with Java, you may have to create a class to print 'Hello World'. But in Python, just a print statement will do. It is also quite easy to learn, understand, and code. Therefore when people pick up Python, they have a hard time adjusting to other more verbose languages like Java.
7. **Readable:** Because it is not such a verbose language, reading Python is much like reading English. This is the reason why it is so easy to learn, understand, and code. It also does not need curly braces to define blocks, and indentation is mandatory. These further aids the readability of the code.
8. **Object-Oriented:** This language supports both the procedural and object-oriented programming paradigms. While functions help us with code reusability, classes and objects let us model the real world. A class allows the encapsulation of data and functions into one.
9. **Free and Open-Source:** Like we said earlier, Python is freely available. But not only can you download Python for free, but we can also download its source code, make changes to it, and even distribute it. It downloads with an extensive collection of libraries to help you with your tasks.

Advantages of Python Over Other Languages

1. Less Coding

Almost all of the tasks done in Python requires less coding when the same task is done in other languages.

Python also has an awesome standard library support, so you don't have to search for any third-party libraries to get your job done. This is the reason that many people suggest learning Python to beginners.

2. Affordable

Python is free therefore individuals, small companies or big organizations can leverage the free available resources to build applications. Python is popular and widely used so it gives you better community support.

3. Python is for Everyone

Python code can run on any machine whether it is Linux, Mac or Windows. Programmers need to learn different languages for different jobs but with Python, you can professionally build web apps, perform data analysis and machine learning, automate things, do web scraping and build games and powerful visualizations.

It is an all-round programming language.

5.3 UML DIAGRAMS

UML is an acronym that stands for Unified Modeling Language. Simply put, UML is a modern approach to modeling and documenting software. In fact, it's one of the most popular business process modeling techniques.

It is based on diagrammatic representations of software components. As the old proverb says: "a picture is worth a thousand words". By using visual representations, we can better understand possible flaws or errors in software or business processes.

UML diagrams, in this case, are used to communicate different aspects and characteristics of a system. However, this is only a top-level view of the system and will most probably not include all the necessary details to execute the project until the very end.

There are several types of UML diagrams and each one of them serves a different purpose. The two broadest categories that encompass all other types are Behavioral UML diagram and Structural UML diagram. As the name suggests, some UML diagrams try to analyze and depict the structure of a system or process, whereas others describe the behavior of the system, its actors, and its building components.

The different types are as follows:

- Use Case Diagram
- Class Diagram
- Sequence Diagram
- Component Diagram

- Collaboration Diagram
- Object Diagram
- State Machine Diagram
- Communication Diagram
- Deployment Diagram

5.3.1 USE CASE DIAGRAM

A use case diagram in the Unified Modeling Language (UML) is a type of behavioral diagram defined by and created from a Use-case analysis. Its purpose is to present a graphical overview of the functionality provided by a system in terms of actors, their goals (represented as use cases), and any dependencies between those use cases. The main purpose of a use case diagram is to show what system functions are performed for which actor. Roles of the actors in the system can be depicted.

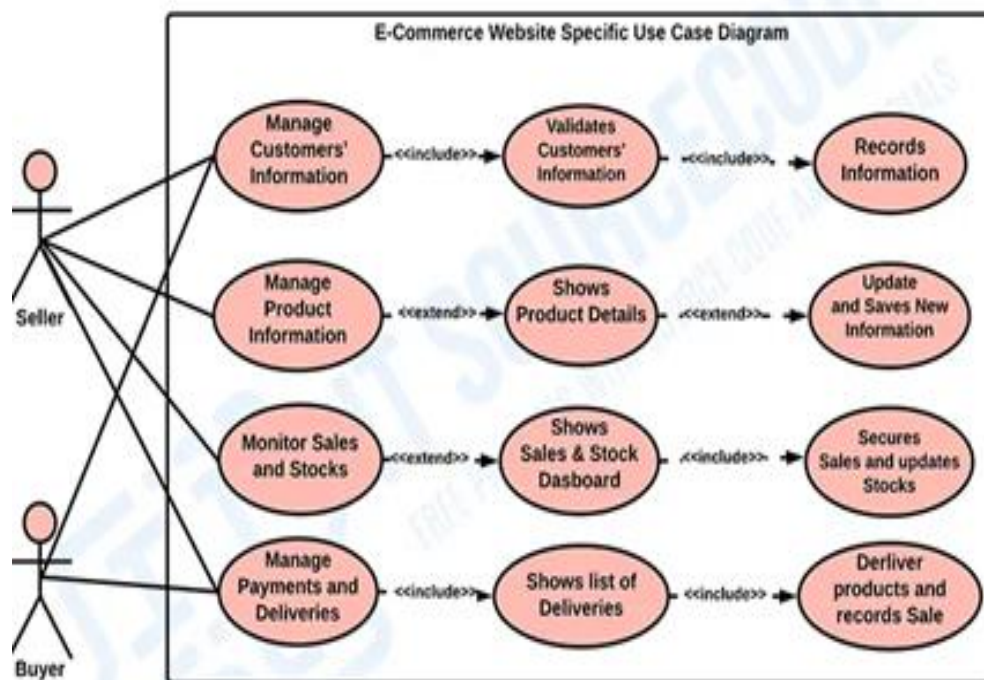


Fig 5.3.1 – Use Case Diagram

5.3.2 CLASS DIAGRAM

Class UML diagram is the most common diagram type for software documentation. Class diagrams contain classes, alongside their attributes (also referred to as data fields) and their behaviors (also referred to as member functions). More specifically, each class has 3 fields: the class name at the top, the class attributes right below the name, the class operations/behaviors at the bottom. The relation between different classes (represented by a connecting line), makes up a class diagram.

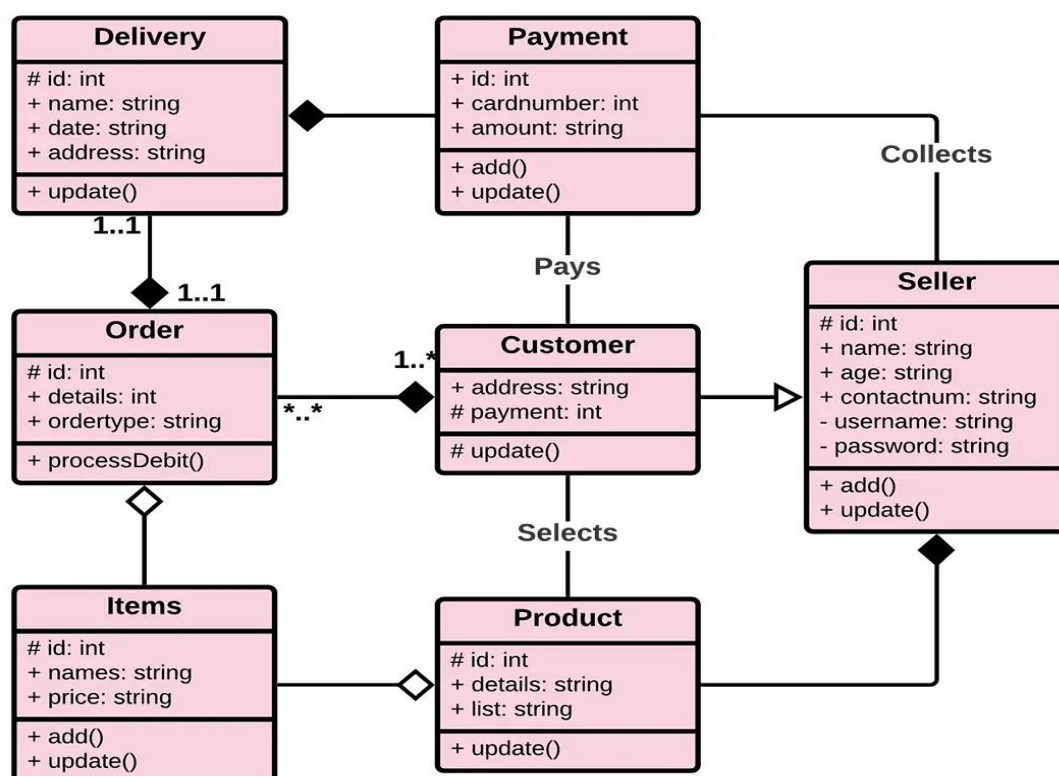


Fig 5.3.2 – Class Diagram

5.3.3 SEQUENCE DIAGRAM

A sequence diagram in Unified Modeling Language (UML) is a kind of interaction diagram that shows how processes operate with one another and in what order. It is a construct of a Message Sequence Chart. Sequence diagrams are sometimes called event diagrams, event scenarios, and timing diagrams.

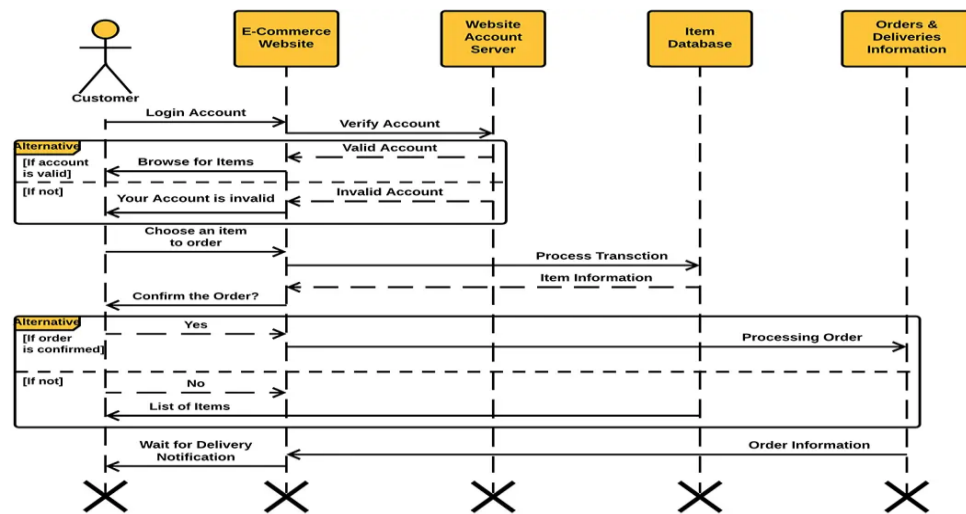


Fig 5.3.3 – Sequence Diagram

5.3.4 ACTIVITY DIAGRAM

Activity diagrams are graphical representations of workflows of stepwise activities and actions with support for choice, iteration and concurrency. In the Unified Modeling Language, activity diagrams can be used to describe the business and operational step-by-step workflows of components in a system. An activity diagram shows the overall flow of control.

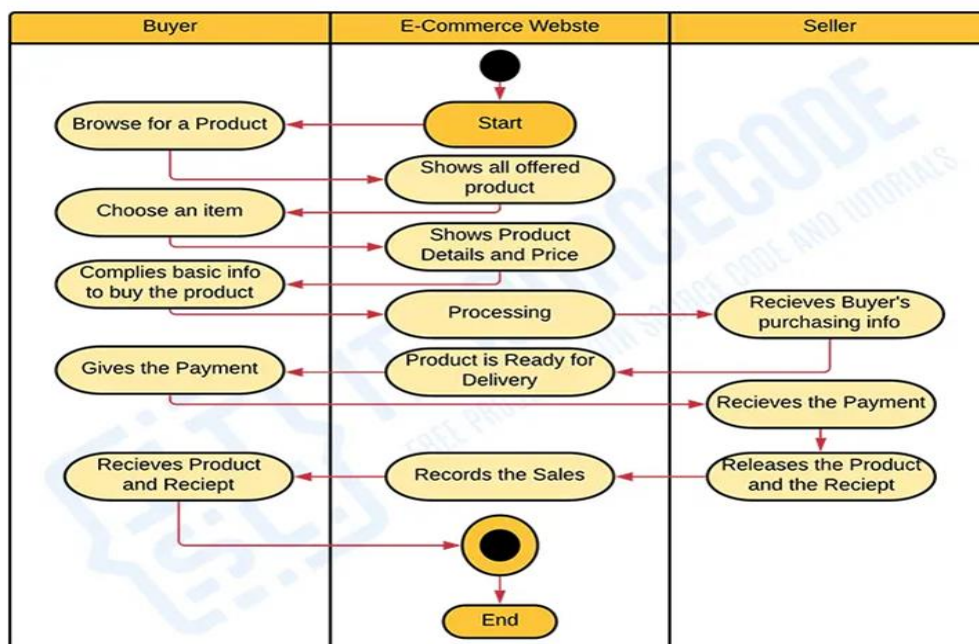


Fig 5.3.4 – Activity Diagram

6. BIBLIOGRAPHY

1. Implisit (Sales Cloud by Salesforce.com). [Online]. Available: <https://www.salesforce.com/blog/2014/08/infographic-7-powerfulpredictors-closed-won-opportunity-gp.html>
2. Insight Squared. [Online]. Available: <https://www.insightsquared.com/features/sales-forecasting/>
3. J. Yan, C. Zhang, H. Zha, et al., "On Machine Learning towards Predictive Sales Pipeline Analytics." Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, pp. 1945-1951, 2015. [Online]. Available: <https://www.aaai.org/ocs/index.php/AAAI/AAAI15/paper/download/9444/9488> [Accessed: Mon. 24 Sept. 2018].
4. M. Bohaneca, M.K. Borstnarb, M. Robnik-Sikonja, "Integration of machine learning insights into organizational learning: A case of B2B sales forecasting." 28th Bled eConference, June 7-10, 2015. [Online]. Available: [https://domino.fov.unimb.si/proceedings.nsf/Proceedings/B12ECF2381AB59EEC1257E5B004B39B7/\\$File/2_Bohanec.pdf](https://domino.fov.unimb.si/proceedings.nsf/Proceedings/B12ECF2381AB59EEC1257E5B004B39B7/$File/2_Bohanec.pdf) [Accessed: Tue. 25 Sept. 2018].
5. M. Bohaneca, M.K. Borstnarb, M Robnik-Sikonja, "Explaining machine learning models in sales predictions." Expert Systems with Applications, no. 71, pp. 416-428, 2017. [Online]. Available: <http://lkm.fri.uni-lj.si/rmarko/papers/Bohanec17-ESwA-preprint.pdf> [Accessed: Tue. 25 Sept. 2018].
6. Gephart, M.A., Marsick, V.J., Mark, E., VanBuren, M.E., Spiro, M.S.: Learning organizations come alive. Training Development 50(12), 36–41 (1996)
7. Nonaka, I., Takeuchi, H.: The Knowledge Creating Organization. Oxford University Press, New York (1995)
8. Senge, P.: The Fifth discipline: The Art & Practice of the Learning Organization. Doubleday Currency, New York (1990)

9. Bohanec, M., Kljajić Borštnar, M., Robnik-Šikonja, M.: Modeling attributes for forecasting B2B opportunities acquisition. In: Proceedings of 34th Conference of Organizational science development, Portorož, Slovenia (2015)
10. Ngai, E.W.T., Xiu, L., Chau, D.C.K.: Application of data mining techniques in CRM: a literature review and classification. *Expert Syst. Appl.* 36, 2592–2602 (2009)
11. Monat, J. P.: Industrial sales lead conversion modeling. *Market. Intell. Plan.* 29(2), 178–194(2011)
12. Rieg, R.: Do forecast improve over time? *Int. J. Account. Inform. Manage.* 18(3) (2010)
13. Alvarado-Valencia, J.A., Barrero, L.H.: Reliance, trust and heuristics in judgmental forecasting. *Comput. Hum. Behav.* 36, 102–113 (2014)
14. Maaß, D., Spruit, M., Waal, P.D.: Improving short-term demand forecasting for short-lifecycle consumer products with data mining techniques, pp. 1–17. *Decision Analytics*, Springer(2014)
15. Witten, I.H., Eibe, F., Hall, M.A.: Data mining—practical machine learning tools and techniques, 3rd edn. Elsevier (2011)
16. Liao, S.H., Chu, P.H., Hsiao, P.Y.: Data mining techniques and applications—a decade review from 2000 to 2011. *Expert Syst. Appl.* 39 (2012)
17. Bose, I., Mahapatra, R.K.: Business data mining—a machine learning perspective. *Inf. Manag.* 39(3), 211–225 (2001)
18. Robnik-Šikonja, M., Kononenko, I.: Explaining classification for individual instances. *IEEE Trans. Knowl. Data Eng.* 20(5), 589–600 (2008)
19. Collopy F., Adya M., Armstrong, J.S.: Expert systems for forecasting. In: *Principles of Forecasting: A Handbook for Researchers and Practitioner*. Kluwer (2001)
20. Sein, M.K., Henfridsson, O., Purao, S., Rossi, M., Lindgreen, R.: Action design research. *MIS Q.* 35, 37–56 (2011)