

MES COLLEGE OF ENGINEERING, KUTTIPPURAM  
DEPARTMENT OF COMPUTER APPLICATIONS  
20MCA245 – MINI PROJECT

---

**PROFORMA FOR THE APPROVAL OF THE THIRD SEMESTER MINI PROJECT**

---

*(Note: All entries of the proforma for approval should be filled up with appropriate and complete information. Incomplete Proforma of approval in any respect will be rejected.)*

Mini Project Proposal No: \_\_\_\_\_  
(Filled by the Department)

Academic Year : 2020-2022  
Year of Admission : 2020

1. Title of the Project : FREQUENT ITEM SET MINING BASED ON DATA SET
2. Name of the Guide : \_\_\_\_\_
3. Number of the Student: MES20MCA-2035
4. Student Details (in BLOCK LETTERS)

Name	Roll Number	Signature
1. <u>NABEEL EP</u>	<u>35</u>	_____

Date: 01/12/2021

---

**Approval Status :**    Approved / Not Approved

Signature of  
Committee Members } \_\_\_\_\_

---

**Comments of The Mini Project Guide**

**Dated Signature**

Initial Submission : \_\_\_\_\_

First Review : \_\_\_\_\_

Second Review : \_\_\_\_\_

---

**Comments of The Project Coordinator**

**Dated Signature**

Initial Submission: \_\_\_\_\_

First Review \_\_\_\_\_

Second Review \_\_\_\_\_

---

Final Comments :

Dated Signature of HOD

# FREQUENT ITEM SET MINING BASED ON DATA SET

NABEEL EP

---

## INTRODUCTION & OBJECTIVE

The rapid generation of data by today's applications has led to a state of data abundance. However, this huge amount of data if left untouched is a significant waste. The data is needed to understand the past and predict the future. According to (Gan et al., 2017), data mining is the solution for this. There are different tasks defined inside data mining. Tasks like classification, clustering, outlier detection, anomaly detection, and frequent pattern mining are examples of typical data mining tasks having great importance each and presenting serious challenges to researchers. As mentioned, frequent pattern mining is one subfield of data mining. It is further deconstructed into subfields such as frequent itemset mining, frequent sequence mining, and frequent graph mining, each having their own set of challenges and respective literature. As generated data accumulates, it turns into big data. Big data is data that is large enough that is not suitable to be fit into a single computer's RAM. Further, big data may not even fit in a single PC's hard drive so that processing it is impossible on such hardware. Therefore, this new phenomenon poses a significant challenge to understanding today's data. To tackle this challenge, traditional approaches to mining and understanding data and also presenting it should be abandoned in favour of new ones which are specifically tailored to this new task. This review is aimed at identifying and summarizing major points about the current effort on big data with a focus on the Apriori algorithm which is an important frequent itemset mining algorithm.

## **OBJECTIVES**

- To identify the item set
- It can be applied in any places in malls or supermarkets etc..

## **Problem definition and initial requirements**

### **EXISTING SYSTEM**

In this existing system the users have to find each products and take the each one so it takes more time to find each products. Persons want to search the items the mostly picked so this system is more time consuming and also more workers are needed for give knowledge about where the each product is so more workers is needed. In the current system if one product is finished most of the people restrict the other one also and move on to other shops that will make a big loss to the company also.

### **PROPOSED SYSTEM**

In this proposed system it is very helpful for all users ie by using data mining easily find most relevant or most frequently bought item set, so based on the frequently bought items create an item set and the users can easily reach and buy the product with minimum time. Based on the item set shops get more profit while a set of product is bought. By using this its most helpful for users and also for shops based on time and also based on profit respectively. Also it the shop needed less staff because these type of item set are placed mostly on front based o item set. So less hand power is needed.

# **Basic functionalities**

## **FUNCTIONAL MODULE**

### **Apriori algorithm**

Apriori is an algorithm for frequent itemset mining and association rule learning over relational databases. It proceeds by identifying the frequent individual items in the database and extending them to larger and larger item sets as those item sets appear sufficiently often in the database.

## **MODULE DESCRIPTION**

- Admin

- User

### **ADMIN**

- Add Products
- View Predicted Itemset
- Add Itemset
- View Rating
- View Transfer

### **USERS**

- wallet
- Update profile
- View Products
- Money transfer
- View Predicted Itemset
- Send rating

# **SOFTWARE AND HARDWARE REQUIREMENTS**

## **HARDWARE REQUIREMENTS**

The selection of hardware is very important in the existence and proper working of any software. Then selection hardware, the size and capacity requirements are also important.

- Processor : Intel Pentium Core i3 and above, 64 bits
- RAM : Min3GB RAM
- HARD DISK: 10 GB

## **SOFTWARE REQUIREMENTS**

One of the most difficult task is selecting software for the system, once the system requirements is found out then we have to determine whether a particular software package fits for those system requirements. The application requirement:

- OPERATING SYSTEM: WINDOWS 10
- FRONT END: HTML, CSS, JAVASCRIPT
- BACK END: Mysql
- IDE USED: JetBrainsPycharm, Android studio
- TECHNOLOGY USED: PYTHON JAVA
- FRAME WORK USED: Flask

MES COLLEGE OF ENGINEERING, KUTTIPPURAM  
DEPARTMENT OF COMPUTER APPLICATIONS  
20MCA245 – MINI PROJECT

---

**PROFORMA FOR THE APPROVAL OF THE THIRD SEMESTER MINI PROJECT**

---

*(Note: All entries of the proforma for approval should be filled up with appropriate and complete information. Incomplete Proforma of approval in any respect will be rejected.)*

Mini Project Proposal No: \_\_\_\_\_  
(Filled by the Department)

Academic Year : 2020-2022  
Year of Admission : 2020

1. Title of the Project : PREDICT YOUR CUSTOMER THROUGH CUSTOMER  
BEHAVIOUR

2. Name of the Guide : \_\_\_\_\_

3. Number of the Student: MES20MCA-2035

4. Student Details (in BLOCK LETTERS)

Name	Roll Number	Signature
1. <u>NABEEL EP</u>	<u>35</u>	_____

Date: 01/12/2021

**Approval Status :** Approved / Not Approved

Signature of  
Committee Members }

---

**Comments of The Mini Project Guide**

Dated Signature

Initial Submission :

First Review :

Second Review :

---

**Comments of The Project Coordinator**

Dated Signature

Initial Submission:

First Review

Second Review

---

Final Comments :

Dated Signature of HOD

# PREDICT YOUR CUSTOMER THROUGH CUSTOMER BEHAVIOUR

NABEEL EP

---

## INTRODUCTION & OBJECTIVE

Customers always play vital role in increasing profit and revenue of every organization; hence, to gain customer satisfaction it is important for the organizational managers to maintain one efficient customer relationship management system by selecting the target customers and maintaining effective relationship with them. Moreover, the CRM system will be helpful for the organization in identifying the most prominent group of customers and their behavior; which will become beneficial for the organization in understanding the retention strategies in a better way. Additionally, higher the customer loyalty, lesser is the customer churn rate; hence using machine learning algorithm such as support vector algorithm can add value in preventing the customer churn. This report will focus on the customer retention with the usage of support vector machine learning in gaining customer loyalty and increasing retention..

## Problem definition and initial requirements

### **EXISTING SYSTEM**

In existing system you need to monitor the people to know if they are satisfied to our service or not, because it's not possible to remember every people also it's not easy to collect all feedbacks from customers and analyse it. So it will decrease your company value if you don't improve the customer service according to customer satisfactions

### **PROPOSED SYSTEM**

In proposed system we provide a better service to customers by understanding the customer demand. We collect feedbacks and details from customer and analyse it properly with the help of machine learning. Powerful churn prediction model will help the organizational management to predict the customer churn. Depending on the complex data of the telecommunication industry, support vector machine can turn out advantageous for predicting the churn rate. The above report has focused on the concept of customer retention along with the churn prediction.

## Basic functionalities

### **FUNCTIONAL MODULE**

#### **• CHURN PREDICTION**

One of the ways to calculate a churn rate is to divide the number of customers lost during a given time interval by the number of active customers at the beginning of the period. For example, if you got 1000 customers and lost 50 last month, then your monthly churn rate is 5 percent.

#### **• MACHINE LEARNING**

Machine learning is a method of data analysis that automates analytical model building. It is a branch of artificial intelligence based on the idea that systems can learn from data, identify patterns and make decisions with minimal human intervention.

## **MODULE DESCRIPTION**

- Shop
- Users
- Staff

## **SHOP**

- View users
- View staff
- Verify staff
- Block/Unblock users
- View feedbacks
- View analysis report
- View complaints and send reply
- Send notification

## **USERS**

- Registration
- View and update profile
- Send feedback
- Send complaint view reply
- View bill report

## **STAFF**

- Registration
- view analysis report
- View notification
- View feedbacks
- View complaints
- Add and manage product
- Generate bill

# **SOFTWARE AND HARDWARE REQUIREMENTS**

## **HARDWARE REQUIREMENTS**

The selection of hardware is very important in the existence and proper working of any software. Then selection hardware, the size and capacity requirements are also important.

- Processor : Intel Pentium Core i3 and above, 64 bits
- RAM : Min3GB RAM
- HARD DISK: 10 GB

## **SOFTWARE REQUIREMENTS**

- OPERATING SYSTEM: WINDOWS 10
- FRONT END: HTML, CSS, JAVASCRIPT
- BACK END: Mysql
- IDE USED: JetBrains Pycharm, Android studio
- TECHNOLOGY USED: PYTHON JAVA
- FRAME WORK USED: Flask



MES COLLEGE OF ENGINEERING, KUTTIPPURAM  
DEPARTMENT OF COMPUTER APPLICATIONS  
20MCA245 – MINI PROJECT

---

**PROFORMA FOR THE APPROVAL OF THE THIRD SEMESTER MINI PROJECT**

---

*(Note: All entries of the proforma for approval should be filled up with appropriate and complete information. Incomplete Proforma of approval in any respect will be rejected.)*

Mini Project Proposal No: \_\_\_\_\_  
(Filled by the Department)

Academic Year : 2020-2022

Year of Admission : 2020

1. Title of the Project : AUTOMATIC DOCUMENT N CLASSIFIER
2. Name of the Guide : \_\_\_\_\_
3. Number of the Student: MES20MCA-2035
4. Student Details (in BLOCK LETTERS)

Name	Roll Number	Signature
1. <u>NABEEL EP</u>	<u>35</u>	_____

Date: 01/12/2021

---

**Approval Status :**    Approved / Not Approved

Signature of  
Committee Members } \_\_\_\_\_

---

**Comments of The Mini Project Guide**

Dated Signature

Initial Submission : \_\_\_\_\_

First Review : \_\_\_\_\_

Second Review : \_\_\_\_\_

---

**Comments of The Project Coordinator**

Dated Signature

Initial Submission: \_\_\_\_\_

First Review \_\_\_\_\_

Second Review \_\_\_\_\_

---

Final Comments :

Dated Signature of HOD

# **AUTOMATIC DOCUMENT CLASSIFIER**

## **NABEEL EP**

---

### **INTRODUCTION & OBJECTIVE**

The Web use has strengthened the creation of digital information in an accelerated way and about multiple topics. The classification of text is widely used to filter e-mails, classify web pages and organize the results recovered by the web browsers. In the process of recovering information, which includes work-tasks of representation, organization, storage and access to the information, it is desired to have associations of documents by keywords at all times. In this way, classifying documents in an automatic way would allow us to find information in a more efficient way

The classification is a grouping procedure which allows us to group a set of data according to a selected criterion. Generally, the objects or data of the same group share similar characteristics with one another, while the objects of different groups will have less similarity among them. For example, in an organization the documents can be classified by a functional criterion, that is to say, grouping the documents by activities inside of the company, or through a criterion of hierarchical order, where the managers have access to different documents that employees have access to. The goal in the task of classification is to locate the document of an appropriate class.

Having a large number of features makes the classification process to be computationally expensive and that the classes not being well defined. The feature selection focuses on reducing the dimensionality, many approaches concentrate on considering only one subset of features extracted from text. Generally, for the selection of characteristics we have techniques based in the collection of documents and techniques based in typifying each class.

### **OBJECTIVES**

The main objective is to :

- Improve the customer experience and throughput rate of your classification -heavy processes without increasing costs.
- Automate the process of grouping documents and use this information to process an entire volume of documents.
- Take documents and easily organize, extract, and apply key metadata to simplify and organize documents into a content management.

### **Problem definition and initial requirements**

#### **EXISTING SYSTEM**

The use of evolutionary algorithms to solve classification problems has been a recurrent approach. The use of the ACM taxonomy was proposed in order to obtain the similarity among documents, where each document is formed by a set of keywords. A methodology to obtain the distance between the words in the ACM taxonomy was designed. Such methodology makes use of the FloydWarshall algorithm, which is typically used to obtain the minimum distance between two nodes in a graph. The grouping of scientific documents was proposed as a problem, for which it was designed a genetic algorithm for classification.

- Uses genetic algorithm for text classification.

- ACM taxonomy is used to measure similarities between documents.
- Floyd Warshall Algorithm used to find distance between nodes in a graph.

## **PROPOSED SYSTEM**

In the proposed system it is expected to evaluate the efficiency of algorithm with higher number of real articles. So, in this way we do the text grouping of document sets of any domain and without being restricted to a few categories. It can be used to decide which document is taken to which cluster, and classification the scientific documents or web documents in a particular cluster.

Additionally, compared the system being developed is capable of classifying documents which belonging to topics which were not provided in the data set into a separate category. It is also capable of identifying duplicate documents present in the input provided if any.

The implementation of this system will be done as a web application and will include User and Admin modules. The admin module will contain some of the main function such as Dataset creation, there are other functions provided such as verifying user, communication with the user etc. The user module contains functions for uploading documents, also includes the functions of sending feedbacks, complaints etc and also view the communications with the admin. The main function of the user module is to view the final classification under the categories, which is the main output of the application.

# **Basic functionalities**

## **FUNCTIONAL MODULE**

### **Genetic Algorithm**

Following is the foundation of GAs based on this analogy –

- Individual in population compete for resources and mate, here the individuals are keywords of the document and of the dataset.
- Those individuals who are successful (fittest) then mate to create more offspring than others.
- Genes from “fittest” parent propagate throughout the generation, that is sometimes parents create offspring which is better than either parent. Here the parent is the Type of the document or the category to which the document is matched.
- Thus each successive generation is expected to have better qualities than previous ones in terms of solutions.

## **MODULE DESCRIPTION**

- Admin
- Student

## **ADMIN**

- Verify the user
- Reply to messages

- Add Datasets

## **STUDENT**

Send and receive Messages.

- Upload Documents.
- View Category.
- Identify duplicate documents.
- View history (All documents which user has uploaded along with its categories).
- Display documents under different categories.

# **SOFTWARE AND HARDWARE REQUIREMENTS**

## **HARDWARE REQUIREMENTS**

The selection of hardware is very important in the existence and proper working of any software. Then selection hardware, the size and capacity requirements are also important.

- Processor : Intel Pentium Core i3 and above, 64 bits
- RAM : Min4GB RAM
- HARD DISK: 10 GB

## **SOFTWARE REQUIREMENTS**

One of the most difficult task is selecting software for the system, once the system requirements is found out then we have to determine whether a particular software package fits for those system requirements. The application requirement:

- OPERATING SYSTEM : WINDOWS 10
- FRONT END : HTML, CSS, JAVASCRIPT
- BACK END : Mysql
- IDE USED : JetBrainsPycharm, Android studio
- TECHNOLOGY USED : PYTHONJAVA
- FRAME WORK USED : Flask

MES COLLEGE OF ENGINEERING, KUTTIPPURAM  
DEPARTMENT OF COMPUTER APPLICATIONS  
20MCA245 – MINI PROJECT

---

**PROFORMA FOR THE APPROVAL OF THE THIRD SEMESTER MINI PROJECT**

---

(Note: All entries of the proforma for approval should be filled up with appropriate and complete information. Incomplete Proforma of approval in any respect will be rejected.)

Mini Project Proposal No: \_\_\_\_\_  
(Filled by the Department)

Academic Year : 2020-2022

Year of Admission : 2020

Title of the Project : **ROBUST DYNAMIC BLOCK BASED IMAGE WATERMARKING IN DWT DOMAIN**

1.

2. Name of the Guide : \_\_\_\_\_

3. Number of the Student: MES20MCA-2035 \_\_\_\_\_

4. Student Details (in BLOCK LETTERS)

Name	Roll Number	Signature
1. <u>NABEEL EP</u>	<u>35</u>	_____

Date: 01/12/2021

**Approval Status :** Approved / Not Approved

Signature of  
Committee Members }

---

**Comments of The Mini Project Guide**

Dated Signature

Initial Submission :

First Review :

Second Review :

---

**Comments of The Project Coordinator**

Dated Signature

Initial Submission:

First Review

Second Review

---

Final Comments :

Dated Signature of HOD

# ROBUST DYNAMIC BLOCK BASED IMAGE WATERMARKING IN DWT DOMAIN

NABEEL EP

## INTRODUCTION & OBJECTIVE

Digital rights management (DRM) is becoming an increasingly important issue in multimedia applications and services. One of the enabling technologies for DRM is digital watermarking. Digital watermarking technique is a solution to resolve copyright protection and information security problems by embedding watermark bits into digital protected media, such as image. They can be also be used for fingerprinting, broadcast monitoring, indexing and medical safety. The watermark bits contain the owner information, which could be the logo, cipher or serial number. When some properties like robustness, security and imperceptibility be observing, then the proposed algorithm can be claim to effective for watermarking purposes. Several algorithms by using different transforms are presented in this domain, like Discrete Cosine Transform (DCT), Fourier Transform (FT), Discrete Wavelet Transform (DWT) and Fractal Transform. The DWT has become the core technique for image processing also image watermarking, because DWT has a number of advantages over other transform such as progressive and low bitrate transmission, quality scalability and region-of-interest (ROI) coding demand more efficient and versatile image coding that can be exploited for both image compression and watermarking applications.

Digital watermarking is a method for protecting copyrighted materials such as digital images. This paper presents a new watermark embedding technique based on Discrete Wavelet Transform (DWT) for hiding little but important information in images. In order to conform to human perception characteristics, this approach uses three subbands of DWT. The main goal of the proposed scheme is to apply a dynamic blocking instead of a static one for selecting the positions of the embedding watermark bits. Dynamic blocking is applied to those pixels which are related to strong edges where these pixels are obtained from HL and LH sub-bands of DWT. Because watermark bits are embedding several times and their embedding position depend to the selected cover image, our proposed scheme is naturally secure to block based local attacks, and therefore, it is suitable for maps and natural images. Robustness of this technique is because of the fact that during the extracting phase, one of these two sub-bands, i.e. HL or LH, has a role of backup for the other one. Experimental results show the high validity and good transparency of this new approach.

## **Problem definition and initial requirements**

### **EXISTING SYSTEM**

Most of the proposed watermarking schemes are nonblock-based schemes, which mean that the same technique is used in all image parts ignoring that each part of the image may have its own characteristics, such as dark areas, texture areas, etc.

- Invisible modification of the least significant bits in the file can be the other way of digital watermarking. Least significant bits are modified in a unique pattern and it is not visible. Layering visible symbol on the top of the image is a method for digital watermarking.
- Using spatial domain methods. These are less complex as no transform is used.

### **PROPOSED SYSTEM**

Transform domain watermarking technique are more robust in comparison to spatial domain methods. Among the transform watermarking technique discrete wavelet transform gaining more popularity. In order to achieve the highest performance in terms of robustness and security, the dynamic block-based watermarking method is used. First, the 3-level Haar wavelet transform is applied on original image, then special blocks that refer to the image edge's pixels are selecting. The mechanism of dynamic blocking in the proposed algorithm is consist that it cluster the variable number of the wavelet coefficients from L<sub>Hn</sub> or H<sub>L</sub><sub>n</sub> subbands into a block related to their grayscale levels, then it quantize the block's coefficients to embed a watermark bits. A general watermarking scheme includes two main steps, a watermark embedding and a watermark extraction. The

proposed algorithm has used the gray image for embedding a pre-defined watermark in to an image and a binary logo image such as watermark. The watermark bit is either 1 or 0. In the embedding process, the original image is first transformed by n-level DWT. There is a tradeoff in the choice of wavelet decomposition levels. It means the parameter n is related to the purpose of embedding, the lower n is using for achieve the better amount of capacity and the higher n is using for better controlling the amount of robustness. The LLnsubband is not suitable for embedding a watermark, because it is a low-frequency subband that contains important information about an image and easily causes image distortion. Also In order to have blind authentication capability, histogram of LLnsubband before and after watermarking should be unchanged. Embedding a watermark in the HHnsubband is also not suitable, since the subband is easily eliminated, for example by lossy compression. Therefore the algorithm utilizes LHn and HLnsubbands for embedding a watermark bit.

## FUNCTIONAL MODULE

### MODULE DESCRIPTION

- Admin
- User

#### ADMIN

**Admin** is the main module of the system .they can control over all sytem.they can view registered user and also block the account of user if any problem occurred.

- Login
- View registered user
- view Feedback
- block user

#### User

**User** is the another module of the system. They user all functionality after the registration is completed.

- Registration
- Login
- Registered user view
- File upload(hidden message)
- Share(view message)
- Send feedback

# **SOFTWARE AND HARDWARE REQUIREMENTS**

## **HARDWARE REQUIREMENTS**

The selection of hardware is very important in the existence and proper working of any software. Then selection hardware, the size and capacity requirements are also important.

- Processor : 64 bit
- RAM : Min 3 GB
- Hard Disk : 10 GB

## **SOFTWARE REQUIREMENTS**

One of the most difficult task is selecting software for the system, once the system requirements is found out then we have to determine whether a particular software package fits for those system requirements. The application requirement:

- OPERATING SYSTEM: WINDOWS 10
- FRONT END: HTML, CSS, JAVASCRIPT
- BACK END: Mysql
- IDE: JetBrainsPycharm, Android studio
- TECHNOLOGY USED: PYTHON, JAVA
- FRAME WORK USED: Flask