**RECOMMENDATION FOR SELECTING SMART VILLAGE IN INDIA THROUGH OPINION MINING**

**A PROJECT REPORT**

*Submitted by*

**NANTHINI . J (810015104052)**

**PAVITHRA . E (810015104058)**

*in partial fulfillment for the award of the degree*

*of*

**BACHELOR OF ENGINEERING**

**IN**

**COMPUTER SCIENCE AND ENGINEERING**



**UNIVERSITY COLLEGE OF ENGINEERING – BIT CAMPUS,**

**TIRUCHIRAPPALLI**

**ANNA UNIVERSITY::CHENNAI 600 025**

**APRIL 2019**

**UNIVERSITY COLLEGE OF ENGINEERING,**

**BIT CAMPUS,**

**TIRUCHIRAPPALLI-620 024**

**BONAFIDE CERTIFICATE**

Certified that this project report “**Recommendation for selecting smart village in India through Opinion mining”** is the bonafide work of **“Ms. J. NANTHINI (810015104052)** and **Ms. E. PAVITHRA (810015104058)”** who carried out the project work under my supervision.

**SIGNATURE SIGNATURE**

**Mr. D. Venkatesan Mr.Prasanna Kumar**

HEAD OF THE DEPARTMENT SUPERVISOR

Assistant Professor Teaching Fellow

Computer Science &Engineering Computer Science &Engineering

University College of Engineering, University College of Engineering,

Anna University-BIT Campus, Anna University-BIT Campus,

Tiruchirappalli-620 024 Tiruchirappalli-620 024

Submitted for the project Viva voce examination held on ………………

**Internal Examiner External Examiner**

**DECLARATION**

We hereby declare the work entitled “**RECOMMENDATION FOR SELECTING SMART VILLAGE IN INDIA THROUGH OPINION MINING”** is submitted in partial fulfillment of the requirement for the award of the degree in B.E., Computer Science and Engineering, University College of Engineering(BIT Campus), Tiruchirappalli, is a record of our own work carried out by us during the academic year 2018-2019 under the supervision and guidance of Mr. M. Prasanna Kumar, Teaching Fellow, Department of Computer Science and Engineering, University College of Engineering(BIT Campus), Tiruchirappalli. The extent and source of information are derived from the existing literature and have been indicated through the dissertation at the appropriate places. The matter embodied in this work is original and has not been submitted for the award of any degree, either in this or any other University.

SIGNATURE OF THE CANDIDATES

J. NANTHINI (810015104052)

E. PAVITHRA (810015104058)

I certify that the declaration made above by the candidate is true.

SIGNATURE OF THE GUIDE

Mr. M. PRASANNAKUMAR

Teaching Fellow,

Department of CSE,

University College of Engineering,

BIT Campus, Anna University,

Tiruchirappalli-620 024.

**ACKNOWLEDGEMENT**

I would like to convey my heartfelt thanks to our honorable Dean **Dr. T. SENTHILKUMAR,** Associate Professor for having provided me with all required facilities to complete my project without hurdles.

I would like to express my sincere thanks and deep sense of gratitude to guide **Mr. D. VENKATESAN,** Assistant Professor and Head, Department of Computer Science and Engineering, for his valuable guidance, suggestions and constant encouragement paved way for the successful completion of this project work.

I would like to thank my project guide **Mr. M. PRASANNAKUMAR,** Teaching Fellow, Department of Computer Science and Engineering, for his valuable guidance throughout the phase of the project. This is our responsibility to thank our project coordinator **Mr. C. SANKAR RAM,** Assistant Professor, Department of Computer science and Engineering for his constant inspiration that he has all through the project period.

I would like to thank **Mr. C. SURESH KUMAR,** Teaching Fellow, Department of Computer Science and Engineering, for his encouragement for this work.

I extend my thanks to all other teaching and non-teaching staffs for their encouragement and support.

I thank my beloved parents and friends, for their full support in my career development of this project.

**ABSTRACT**

In our country, most of the people are staying in below poverty line. Nowadays, village people are also most inadequate with mobile phones. To develop this village as smart, to emphasize on different factors like agriculture, employment, nutrition security, environment, natural resource utilization, and conservation, etc. To select smart village based on collection of different opinion from village people in terms of forms, questionnaire, views and surveys, etc. Opinion mining extracts useful knowledge about village from ample of opinions. This mining process gives us a right direction for creating smart village. In this work, we are trying to create digitalized village which is basically an application of Information and Communications Technology to deﬁne the major function of Government in order to bring about Small, Moral, Accurate, Reliable and Transparent. Opinion mining extracts variety opinions from villagers by above methods. To know how much response from villagers in different criteria for smart village, we use datamining techniques which uses classification and prediction techniques for calculating number of person’s opinion.

|  |  |  |
| --- | --- | --- |
| **CHAPTER**  **NO.** | **TITLE**  **ABSTRACT**  **LIST OF FIGURES**  **LIST OF ABBREVIATIONS** | **PAGE NO**.  v  ix  x |
| 1 | **INTRODUCTION** | 1 |
|  | 1.1 Overview of Smart Village | 1 |
|  | 1.2 Recommendation System | 2 |
|  | 1.3 Sentiment Analysis | 4 |
|  | 1.4 Data Collection | 6 |
|  | 1.5 Opinion Mining | 6 |
|  | 1.6 Opinion Mining Algorithms | 8 |
| 2 | **LITERATURE SURVEY** | 11 |
| 3 | **SYSTEM ANALYSIS** | 21 |
|  | 3.1 Existing System | 21 |
|  | 3.1.1 Limitations | 22 |
|  | 3.2 Proposed System | 22 |
| 4 | **SYSTEM SPECIFICATION** | 23 |
|  | 4.1 Software Requirements | 23 |
|  | 4.2 Hardware Requirements | 23 |
|  | 4.3 About the Software | 24 |
| 5 | **SYSTEM DESIGN** | 27 |
|  | 5.1 System Architecture | 27 |
|  | 5.2 Module Description | 28 |
|  | 5.2.1 Data Collection | 28 |
|  | 5.2.2 Data Preprocessing | 30 |
|  | 5.2.3 SVM | 32 |
|  | 5.2.4 Response Detection  5.2.5 Recommendation System | 33  33 |
| 6 | **CONCLUSION AND FUTURE WORKS** | 34 |
| **APPENDICES** |  |  |
|  | Appendix1 Sample Source Code | 35 |
|  | Appendix2 Sample Screenshots | 39 |
| **REFERENCES** |  | 44 |

**TABLE OF CONTENTS**

**LIST OF FIGURES**

|  |  |  |
| --- | --- | --- |
| **FIGURE NO.** | **FIGURE NAME** | **PAGE NO.** |
| 5.1 | Architecture of the Proposed System | 24 |
| 5.2 | Architecture of Data Preprocessing | 30 |

|  |  |
| --- | --- |
|  |  |

|  |  |  |
| --- | --- | --- |
|  | **ABBREVIATIONS**  API Application Programming Interface  CSV Comma Separated Values  SVM Support Vector Machine  SAGY Sanad Adarsh Gram Yojana  ATVT Apno Taluko Vibrant Taluko  DBT Direct Benefit Transfer  IDS Intrusion Detection System  GUI Graphical User Interface |  |
|  | MAP Maximum A Posteriori  HDFS Hadoop Distributed File System  SMAC Social media Mobile Analytics and Cloud  NLP Natural Language Processing  CSV Comma Separated Values |  |
|  |  |  |

**CHAPTER 1**

**INTRODUCTION**

**1.1 OVERVIEW OF SMART VILLAGE**

A village is a clustered human settlement or community, larger than a hamlet but smaller than a town, with a population ranging from a few hundred to a few thousand. Though villages are often located in rural areas, the term urban village is also applied to certain urban neighbourhoods. Smart Village is a concept adopted by national, state and local governments of India, as an initiative focused on holistic rural development, derived from Mahatma Gandhi's vision of Adarsh Gram (Ideal Village) and Swaraj (Self Reliance). Prime Minister Narendra Modi launched Sanad Adarsh Gram Yojana (SAGY) or SAANJHI) on 2nd October 2014, Gandhi's birthday, in addition to Smart Cities" and "Digital India", as a development program for India. Smart Village means a village which is aware to increase facilities for the citizen by taking decisions democratically and Smart Village means a village in which the youth, women, farmers, village artisans, backward and deprived people may get equal opportunity for development.

In Smart Villages access to sustainable energy services acts as a catalyst for development – enabling the provision of good education and healthcare, access to clean water, sanitation and nutrition, the growth of productive enterprises to boost incomes, and enhanced security, gender equality and democratic engagement. The concept of the smart village is not constructed on the image of a city or a very developed village of some states or nations. It is not an adopted smartness, but an avenue to show the inherent smartness of the villages.

**1.2 OVERVIEW OF RECOMMENDATION SYSTEM**

A recommender team or a recommendation system (sometimes replacing "system" with a synonym such as platform or engine) is a subclass of information filtering system that seeks to predict the "rating" or "preference" a user would give to an item. Recommender systems are utilized in a variety of areas including movies, music, news, books, research articles, search queries, social tags, and products in general. For the development of smart village, a cluster approach should be followed. To develop the culture of that village, we should look positively into some parameters such as education, agriculture, environment, road development, natural resource management like water preservation and soil preservation, etc. Each district can be divided into clusters of 10 to 15 villages according to the specific number of populations.  There is need to design the blue print to develop infrastructure like roads, buildings, bridges, canals, ponds, sewages, schools, colleges, hospitals etc. as per need based on local resources available or in the district or within the state. The smart village concept should formulate growth strategies for the village to make it self-sufficient in protecting native occupation and heritage and monuments of the village. Pure drinking water and good sanitation are essential prerequisites for good health and hygiene.

Most of the epidemics and ill health in India is mainly due to communicable diseases caused by oral fecal routes. Solar powered street lights have all the required means today to lighten up the villagers in terms of the sense of security. Solar LED street lighting will provide a high quality, sustainable lighting solution for people in remote areas who don’t have access to the conventional electricity grid. It will help in increasing the level of safety on roads and streets and allowing for more economic and social activity. Unorganized growth model has seen in the forms of mammoth heaps of waste in many cities.

**Mandatory targets for Smart Village**

Following mandatory targets are necessary to achieve before declaring any Gram Panchayat a Smart Village.

* 100% vaccination
* 0% drop out ratio
* Open defecation free village
* 90% tax collection
* Compulsory door to door solid waste disposal
* Distribution of ATVT services at E-Gram Center
* ‘A’ grade primary school
* 100% individual toilet
* Remove encroachments on public roads
* Malnutrition - No child should be in ‘RED Zone’
* Mandatory organizing four Gram Sabha in a year with 50% presence
* 100% implementation of direct deposit of Government assistance into beneficiary’s account. (D.B.T.)
* 100% enrollment of children in Anganwadi.
* To update revenue/ Gram Panchayat records.
* Road side plantation in the village.
* 100% Wi-Fi village.

**Social Development**

* Violence and crime free village
* 100% literate village
* Malnutrition free village
* Child-marriage free village
* Pollution free village

**Human Development**

* 100% coverage of widow, disabled and senior citizens for assistance
* 100% girls’ education up to standard 8
* Reduction in child mortality rate
* Qualitative education

**Economic Sector**

* Awareness toward organic farming/ vermicompost method
* Awareness towards modern farming and drip irrigation
* Development of dairy industry
* Opportunities of employment

**Community facilities**

* Availability of pure drinking water to each household
* Door-to-door solid waste disposal in each area of the village
* Pasture and public garden development

**1.3 OVERVIEW OF SENTIMENT ANALYSIS**

Opinion mining refers to the use of natural language processing, text analysis, computational linguistics, and biometrics to systematically identify, extract, quantify, and study affective states and subjective information. Sentiment analysis is widely applied to voice of the customer materials such as reviews and survey responses, online, social media, and healthcare materials for applications that range from marketing to customer service to clinical medicine. Sentiment analysis of natural language texts is a large and growing field. Opinion Mining is the computational treatment of opinions, sentiments and subjectivity of text. Sentiment analysis is a Natural Language Processing and Information Extraction task that aims to obtain writer’s feelings expressed in positive or negative comments, questions and requests, by analyzing a large number of documents. Converting a piece of text to a feature vector is the basic step in any data driven approach to Sentiment analysis. Term frequency has always been considered essential in traditional Information Retrieval and Text Classification tasks.

**Techniques Behind Sentiment Analysis**

The following techniques used for sentiment analysis

**Sentiment Classification**

It is used to analysis specific sentences or words combination to generate a number telling massager's mood from Tweets messages by identified users. There are three main classification levels in SA:

1. Document level
2. Sentence level
3. Aspect level

**Document Level:** Document-level SA aims to classify an opinion document as expressing a positive or negative opinion or sentiment. It considers the whole document a basic information unit (talking about one topic).

**Sentence Level:** Sentence-level SA aims to classify sentiment expressed in each sentence. The first step is to identify whether the sentence is subjective or objective. If the sentence is subjective, Sentence-level SA will determine whether the sentence expresses positive or negative opinions.

**Aspect Level:** Aspect-level SA aims to classify the sentiment with respect to the specific aspects of entities. The first step is to identify the entities and their aspects. The opinion holders can give different opinions for different aspects of the same entity like this sentence “The voice quality of this phone is not good, but the battery life is long”.

**1.4 DATA COLLECTION**

Data collection is the process of gathering and measuring information on variables of interest, in an established systematic fashion that enables one to answer stated research questions, test hypotheses, and evaluate outcomes. collection of different opinion from village people in terms of forms, questionnaire, views and surveys, etc.

**Data Preprocessing:** Data preprocessing is an important step in the data mining process. The phrase "garbage in, garbage out" is particularly applicable to data mining and machine learning projects. Data-gathering methods are often loosely controlled, resulting in out-of-range values (e.g., Income: −100), impossible data combinations (e.g., Sex: Male, Pregnant: Yes), missing values, etc. Analyzing data that has not been carefully screened for such problems can produce misleading results.

**Data Transformation:** Data transformation is the process of converting data from one format or structure into another format or structure. It is a fundamental aspect of most data integration and data management tasks such as data wrangling, data warehousing, data integration and application integration.

**Data Reduction:** Data reduction is the transformation of numerical or alphabetical digital information derived empirically or experimentally into a corrected, ordered, and simplified form. The basic concept is the reduction of multitudinous amounts of data down to the meaningful parts.

**1.5 OPINION MINING**

It also known as Sentiment analysis refers to the use of natural language processing, text analysis and computational linguistics to identify and extract subjective information in source materials. Textual information in the world can be broadly classified into two main categories, facts and opinions.

Facts are objective statements about entities and events in the world. Opinions are subjective statements that reflect people’s sentiments or perceptions about the entities and events. Much of the existing research on text information processing has been (almost exclusively) focused on mining and retrieval of factual information, e.g., information retrieval, Web search, and many other text mining and natural language processing tasks. This is not only true for individuals but also true for organizations. Sentiment analysis is widely applied to reviews and social media for a variety of applications, ranging from marketing to customer service.

The aim of sentiment analysis is to determine the attitude of a speaker or a writer with respect to some topic or the overall contextual polarity of a document. The attitude may be his or her judgment or evaluation, affective state i.e. the emotional state of the author when writing, or the intended emotional communication i.e. the emotional effect the author wishes to have on the reader. Humans like to express their opinions and are eager to know others' opinions. Automatically mining and organizing opinions from heterogeneous information sources are very useful for individuals, organizations and even governments. Opinion extraction, opinion summarization and opinion tracking are three important techniques for understanding opinions.

**Opinion Extraction**: Its identifying components which express opinions is fundamental for summarization, tracking, and so on. The goal of opinion extraction is to detect where in documents opinions are embedded. Opinions are hidden in words, sentences and documents. An opinion sentence is the smallest complete semantic unit from which opinions can be extracted.

**Opinion Outline:**  It also called a hierarchical outline, is a list arranged to show hierarchical relationships and is a type of tree structure. An outline is used to present the main points (in sentences) or topics (terms) of a given subject. Each item in an outline may be divided into additional sub-items. If an organizational level in an outline is to be sub-divided, it shall have at least two subcategories, as advised by major style manuals in current use. An outline may be used as a drafting tool of a document, or as a summary of the content of a document or of the knowledge in an entire field.

**Response Detection:** An intrusion detection system (IDS) is a device or software application that monitors a network. Systems with response capabilities are typically referred to as an intrusion prevention system. Intrusion detection systems can also serve specific response.

**1.6 OPINION MINING ALGORITHMS**

**K-nearest neighbor**

K-Nearest Neighbor algorithm that is being widely used for classification and regression and also it is a non-parametric method. Every training set that is being present in the multidimensional feature space are the vectors with the specific class labels specified. In n-dimensional space each attribute will be pointing to the training samples with n dimensional numeric attributes. The training phase of the algorithm it acts by storing the feature vectors and class labels. When an unknown sample is given to k-nearest neighbor algorithm it searches for the pattern space for the k training samples that are closer to the unknown samples. Euclidean distance determines the property of the “closeness” measures. When KNN approach is to be applied value should be appropriate and the effectiveness of this approach mainly depends upon this value. It can be widely adopted for multi-class model classes and also for the objects with multiple class labels. KNN is an efficient classification algorithm that is easy for understanding. It is robust even in the case of large dataset with noisy training data being used.

**Naive Bayes classification**

A Naive Bayes classifier is a probabilistic machine learning model that’s used for classification task. The crux of the classifier is based on the Bayes theorem. One of the easiest ways of selecting the most probable hypothesis given the data that we have that we can use as our prior knowledge about the problem. Bayes’ Theorem provides a way that we can calculate the probability of a hypothesis given our prior knowledge. After calculating the posterior probability for a number of different hypotheses, you can select the hypothesis with the highest probability. This is the maximum probable hypothesis and may formally be called the Maximum a Posteriori (MAP) hypothesis. naive Bayes is a classification algorithm for binary (two-class) and multi-class classification problems. The technique is easiest to understand when described using binary or categorical input values. It is called naive Bayes or idiot Bayes because the calculation of the probabilities for each hypothesis are simplified to make their calculation tractable. Rather than attempting to calculate the values of each attribute value, they are assumed to be conditionally independent given the target value and calculated. This is a very strong assumption that is most unlikely in real data, i.e. that the attributes do not interact. Nevertheless, the approach performs surprisingly well on data where this assumption does not hold.

**Support Vector Machine**

Support Vector Machines (SVMs) are supervised learning methods used for classification and regression tasks that originated from statistical learning theory. As a classification method, SVM is a global classification model that generates non-overlapping partitions and usually employs all attributes. The entity space is partitioned in a single pass, so that flat and linear partitions are generated. SVMs are based on maximum margin linear discriminants, and are similar to probabilistic approaches, but do not consider the dependencies among attributes. Traditional Neural Network approaches have suffered difficulties with generalization, producing models which overfit the data as a consequence of the optimization algorithms used for parameter selection and the statistical measures used to select the best model. SVMs have been gaining popularity due to many attractive features and promising empirical performance. A classification task usually involves training and test sets which consist of data instances. Each instance in the training set contains one target value (class label) and several attributes (features). The goal of a classifier is to produce a model able to predict target values of data instances in the testing set, for which only the attributes are known.

**Maximum entropy**

The Max Entropy classifier is a probabilistic classifier which belongs to the class of exponential models. Unlike the Naive Bayes classifier that we discussed in the previous article, the Max Entropy does not assume that the features are conditionally independent of each other. The Max Entropy classifier can be used to solve a large variety of text classification problems such as language detection, topic classification, sentiment analysis and more. Due to the minimum assumptions that the Maximum Entropy classifier makes, we regularly use it when we don’t know anything about the prior distributions and when it is unsafe to make any such assumptions. Moreover, Maximum Entropy classifier is used when we can’t assume the conditional independence of the features. This is particularly true in Text Classification problems where our features are usually words which obviously are not independent. The Max Entropy requires more time to train comparing to Naive Bayes primarily due to the optimization problem that needs to be solved in order to estimate the parameters of the model. Nevertheless, after computing these parameters, the method provides robust results and it is competitive in terms of CPU and memory consumption.

**CHAPTER 2**

**LITERATURE SURVEY**

In this section the literature survey has been carried out. The main focus is on opinion mining, polarity detection techniques, analyzing human emotions, opinion extraction, summarizing the reviews, etc. The literature survey gives the clear idea for proposed system.

Kumar. P, Zaidi. A, et.al [1] presented “Evolving sentiments towards e-governance using opinion mining”. They have considered E-Governance basically as the application of Information and Communications Technology to define the functioning of Government in order to bring about “Simple, Moral, Accountable, Responsive and Transparent” (SMART) governance. In their work, they are trying to evolve the sentiments towards E-Government projects and policies on the basis of opinion collected from the people. They try to analyze the effectiveness of E-governance using the concept of opinion mining. For that, they collected the opinions of people regarding smart city.

Sahoo. A.K, Tiwari. M et.al [2] presented “Signature based malware detection for unstructured data in Hadoop”. They considered about the Software Defined Networks is an emerging network paradigm which introduces programmability to networks and has the capability to dynamically configure the network. SDN takes another approach by decoupling the controller part from the data plane part. It is the next generation Internet technology that solves the ossification of the Internet, along with creating massive innovations so that network management can be easily handled. In that process, the researchers trying to explore the different layers, various protocols supported by SDN, along with discuss the numerous benefits of it. Here opinion mining can be applied on villagers to get their opinions and views. These views collected from people are normally made in an unstructured way. To make it as structured, they used the map reduce approach which can be applied on a huge amount of collected data called big data from people in Hadoop environment. Hadoop is an open-source framework that allows to store and process big data in a distributed environment using map reduce approach. Hadoop Distributed File System (HDFS) is associated with Hadoop which is purely based on Java. Hadoop automatically indexes the file, breaks the file into blocks and stores it in different nodes. After storing in HDFS, it requires map reduce approach for processing.

Jaskaran Kaur, Sheveta Vashisht [3] presented “Analysis and Identifying Variation in Human Emotion through Data Mining”, is explored to detect emotion variation in adolescent age group through data mining techniques. In this age mood swings are maximum as compared to other age group. Children become moodier and more irritable during this period of their lives. This moodiness is commonly attributes to the sudden and fluctuating hormonal level, brain still developing, peer pressure, Cognitive immaturity. Some people get more hyper but others feel under the weather and just want to be left alone, but everyone is different, and it isn't always the same emotion change. During this age period the people around us will experience drastic changes in our moods. If you're out-of-control happy one minute, totally depressed the next minute, and furious the next, you're completely normal! But mood swings can also be frightening and confusing. That research is based on collection of datasets from this large group of emotions in adolescent age by using data mining technique. In their work they tried to propose sentiment analysis by classifying emotions and using decision tree different emotional variations are analyzed. Outlier analysis is used to identify emotion variation in child having any kind of disability.

A. Jeyapriya, C.S. Kanimozhi Selvi [4] presented “Extracting Aspects and Mining Opinions in Product Reviews Using Supervised Learning Algorithm”, they proposed the Social media is emerging rapidly on the internet. This media knowledge helps people, company and organizations to analyze information for important decision making. Opinion mining is also called as sentiment analysis which involves in building a system to gather and examine opinions about the product made in reviews or tweets, comments, blog posts on the web. Sentiment is classified automatically for important applications such as opinion mining and summarization. To make valuable decisions in marketing analysis where implement sentiment classification efficiently. Reviews contain sentiment which is expressed in a different way in different domains and it is costly to annotate data for each new domain. The analysis of online customer reviews in which firms cannot discover what exactly people liked and did not like in document-level and sentence-level opinion mining. It is used to extract most important aspects of an item and to predict the orientation of each aspect from the item reviews.

 Hu Minqing, Liu Bing [5] presented "Mining and summarizing customer reviews",In Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*,* they proposed different techniques for Merchants selling products on the Web often ask their customers to review the products that they have purchased and the associated services For a popular product, the number of reviews can be in hundreds or even thousands. This makes it difficult for a potential customer to read them to make an informed decision on whether to purchase the product. It also makes it difficult for the manufacturer of the product to keep track and to manage customer opinions. In their research, they aim to mine and to summarize all the customer reviews of a product. This summarization task is different from traditional text summarization because they only mine the features of the product on which the customers have expressed their opinions and whether the opinions are positive or negative.

Eirinaki Magdalini, Shamita Pisal, et.al [6] presented "Feature-based opinion mining and ranking", they proposed the proliferation of blogs and social networks presents a new set of challenges and opportunities in the way information is searched and retrieved. Even though facts still play a very important role when information is sought on a topic, opinions have become increasingly important as well. Opinions expressed in blogs and social networks are playing an important role influencing everything from the products people buy to the presidential candidate they support. They present an algorithm which not only analyses the overall sentiment of a document/review, but also identifies the semantic orientation of specific components of the review that lead to a particular sentiment. The algorithm is integrated in an opinion search engine which presents results to a query along with their overall tone and a summary of sentiments of the most important features.

Nayak. A.S, Fernandes. R [7] presented “A study and analysis of opinion mining and sentiment analysis”. In their work they considered on the outbursts of social media has created millions of opportunities for people or web users to state or express their opinions, but it is a challenging task when it comes to bringing out the meaning or some sense from the opinions. Opinion mining can be employed in many ways depending on domain or user interests in different fields for different purpose, due to which analyzing the opinions has become very essential. Opinion mining is a process of extracting the people’s views about a particular topic under discussion, organization and product. These views are collected through social media and review sites like blogs, comments, tweets or reviews. Using opinion mining concept, they extracted valuable factors which will be taken as main criteria of smart village from villagers.

Vidisha M. Pradhan, et al [8] presented “A Survey on Sentiment Analysis Algorithms for Opinion Mining”, Opinion mining and sentiment analysis is rapidly growing area. There are numerous e-commerce sites available on internet which provides options to users to give feedback about specific product. These feedbacks are very much helpful to both the individuals, who are willing to buy that product and the organizations. An accurate method for predicting sentiments could enable us, to extract opinions from the internet and predict customer’s preferences. Before applying any algorithm for polarity detection, pre-processing on feedback is carried out. From these pre-processed reviews opinion words and object on which opinion is generated are extracted and any opinion mining technique is applied to find the polarity of the reviews. In their work, they search various algorithms for sentiment analysis are and challenges and applications appeared in that field.

Anu Peisker and Soumya Dalai [9] presented “Data Analytics for Rural Development”, the impact of big data is enormous virtually in every business sector. The usage so far has been more focused on the e-commerce and marketing sectors. But the wide reach of big data can provide much more innovative, profitable and yet beneficial solutions for many perennial problems faced by different sectors. Rural India plays a vital role in economic growth of the country through agriculture, self-employment, construction, services, etc. The developments in telecommunication sector provide suitable platform for applying data analytics in rural India. The focus of the Government to build digital India through broadband highways connecting every household, village, panchayat, Government department will generate huge amount of data which can be analyzed to provide solution to the never-ending problems of rural India and to create smarter villages. In their work they provide conceptual framework for the application of data analytics in enhancing rural development by supporting different sectors such as agriculture, banking, governance and healthcare.

John DALE, Sunil ISHAIRZAY [10] presented “from Smart Cities to Smart Villages: New Sustainable Futures for Disrupting Rural Migration in Myanmar and India”. Smart cities represent the new engines of growth in the global development of knowledge capitalism.  Europe and the United States generally celebrate this rapidly emerging mode of urban development for its capacity to enhance more sustainable environmental practices, extend the institutional reach of democratic governance, and improve the efficiency of urban management.  Increasingly, smart city development has been encouraged by international development agencies like USAID and DFID as the paradigm of choice for addressing the social problems they anticipate in the emerging “megacities” of the world.  Large corporations are also finding profitable niches in the global business of smart city development. Qatar-based Ooredoo, for example, now controls a growing portfolio of ten Asian megacities which represent its significant investment in smart city development. It seeks specifically to institutionalize a different future for rural villages that reverses the new patterns of rural to urban migration, particularly among young adults, that smart city development has amplified in the region.

ByIon SMEUREANU, Cristian BUCUR [11] presented “Applying Supervised Opinion Mining Techniques on Online User Reviews” is the expression of opinions of users in specialized sites for evaluation of products and services, and also on social networking platforms, has become one of the main ways of communication, due to spectacular development of web environment in recent years. The large amount of information on these platforms make them viable for use as data sources, in applications based on opinion mining and sentiment analysis. The research presented a method of sentiment analysis, on the review made by users to movies. Classification of reviews in both positive and negative classes is done based on a naive Bayes algorithm. As training data, they used a collection (pre-classified in positive and negative) of sentences taken from the movie reviews. To improve classification, they removed insignificant words and introduced in classification groups of words (n-grams). For n = 2 groups we achieved a substantial improvement in classification. As an extension of the research presented in the research, they want to improve the algorithm, enriching the training set of examples, on the way, with examples classified as strong positive or negative, by an established score of classification. They tried to determine, in a review, those sentences which do not express opinions, or determine opinions about the film or the film actors and identify opinions addressed strictly on these items. In their work they tried to highlight the main aspects on which opinions are expressed and to extract opinions based on aspects identification.

Akshi Kumar and Abhilasha Sharma [12] presented “Systematic Literature Review on Opinion Mining of Big Data for Government Intelligence”, With the advent of new technology paradigm, SMAC (Social media, Mobile, Analytics and Cloud) the information network generates an infinite ocean of data spreading faster and larger than earlier. High quality information extracted from this massive volume of data, named as big data, urges the development of an efficient and effective decision support system and powerful strategic tools in the area of government intelligence. Digitally intelligent and smart governance has been identified as a dynamic field with new studies being reported at various research avenues. The need of review, analyze and evaluate research studies across literature is thus fostered motivating us to identify existing trends, research gaps and potential directions of future work within this domain. That research intends to provide a systematic literature review within the promising area of opinion mining and its application to the area of government intelligence.

Karina Gibert, Miquel Sanchez–Marre, et.al [13] presented “A survey on pre-processing techniques: Relevant issues in the context of environmental data mining” is one of the important issues related with all types of data analysis, either statistical data analysis, machine learning, data mining, data science or whatever form of data-driven modeling, is data quality. The more complex the reality to be analyzed is, the higher the risk of getting low quality data. Unfortunately, real data often contain noise, uncertainty, errors, redundancies or even irrelevant information. Useless models will be obtained when built over incorrect or incomplete data. However, pre-processing has not been properly systematized yet, and little research is focused on this. In that research a survey on most popular pre-processing steps required in environmental data analysis is presented, together with a proposal to systematize it. Rather than providing technical details on specific pre-processing techniques, the work focuses on pro viding general ideas to a non-expert user, who, after reading them, can decide which one is the more suitable technique required to solve his/her problem.

Michael D. Ekstrand et.al [14] presented “Collaborative Filtering Recommender Systems” they proposed the recommender systems are an important part of the information and e-commerce ecosystem. Specific tasks, information needs, and item domains represent unique problems for recommenders, and design and evaluation of recommenders needs to be done based on the user tasks to be supported. Effective deployments must begin with careful analysis of prospective users and their goals. Based on this analysis, system designers have a host of options for the choice of algorithm and for its embedding in the surrounding user experience. Their research discusses a wide variety of the choices available and their implications, aiming to provide both practitioners and researchers with an introduction to the important issues underlying recommenders and current best practices for addressing these issues.

Richa Sharma, Shweta Nigam, et.al [15] presented “Polarity Detection at Sentence Level”, they considered on Opinion Mining or Sentiment Analysis is the mining of attitudes, opinions, and emotions automatically from text, speech, and database sources through Natural Language Processing (NLP). In their research an opinion mining system is proposed using unsupervised technique to determine the polarity of sentences to classify the sentences as positive, negative or neutral. Negation is also handled in the proposed system.

**CHAPTER 3**

**SYSTEM ANALYSIS**

**3.1 EXISTING SYSTEM**

* In existing system, extracting aspects and mining opinion process done by the Frequent itemset mining, Naive Bayes algorithm. Using these algorithms, they reach 92 percentage accuracy for the performance.
* Taking different opinions from villagers, then we use those Map Reduce approach on collected unstructured data set to get right conclusion, here we use opinion mining concept and map reduce approach.
* Using Opinion mining, natural language processing and knowledge extraction to Collect the reviews through social media and complete those process.
* A novel method for twitter sentiment mining used by framework-based learner’s emotional state classification and visualization for E-learning systems.
* Opinion extraction from the web, surveys and other which are available digitally using Frequent item set mining and Naive Bayes supervised Learning Algorithm.
* To Get Different emotions by constructing the emotional lexicon and identifying different linguistic features is based on Bayesian probability.
* To determine the polarity of a large number of customer reviews for the products using by Lexicon dictionary-based approach.
* To Analysis and Identifying Variation in Human Emotion Through Data Mining by using decision tree algorithm.

**3.2 LIMITATIONS**

* Performance of sentiment analysis is not efficient.
* Error in statistical measure
* Incomplete training data and continuous variables doesn’t work.
* Accuracy rate is not good.

**3.3 PROPOSED SYSTEM**

In India, most of the people are staying in below poverty line. To develop this village as smart, we emphasize on different factors like agriculture, employment, nutrition security, sanitation, environment, natural resource utilization, and conservation, transport etc. To select smart village based on collection of different opinion from village people in terms of questionnaire and surveys through web application. Opinion mining extracts useful knowledge about village from ample of opinions. This mining process gives us a right direction for creating smart village. In this work, we are trying to create digitalized village which is basically an application of Information and Communications Technology to define the major function of Government in order to bring about Small, Moral, Accurate, Reliable and Transparent. To get accurate response for the performance by using support vector machine algorithm.

**CHAPTER 4**

**SYSTEM SPECIFICATION**

**4.1 SOFTWARE REQUIREMENTS**

Operating System : Windows Pro

Front End : Anaconda Navigator

Back End : MySQL

Language : Python

**4.2 HARDWARE REQUIREMENTS**

Processor : Intel Core i3

Hard Disk : 512 GB

RAM : 4 GB

* 1. **ABOUT THE SOFTWARE**

**Anaconda Navigator**

Anaconda Navigator is a desktop graphical user interface (GUI) included in Anaconda® distribution that allows you to launch applications and easily manage conda packages, environments and channels without using command-line commands. Navigator can search for packages on Anaconda Cloud or in a local Anaconda Repository. Jupyter Notebooks are an increasingly popular system that combine your code, descriptive text, output, images and interactive interfaces into a single notebook file that is edited, viewed. It is available for Windows, macOS and Linux. Anaconda® is a package manager, an environment manager, a Python/R data science distribution, and a collection of over 1500+ open source packages. Anaconda is free and easy to install, and it offers free community support. Easily deploy your projects into interactive data applications, live notebooks and machine learning models with APIs. Share those applications with colleagues and collaborators. Manage your data science assets: notebooks, packages, environments and projects in an integrated data science experience.Navigator is an easy, point-and-click way to work with packages and environments without needing to type conda commands in a terminal window.

**Conda**

Conda is an open source package management system and environment management system that runs on Windows, macOS and Linux. Conda quickly installs, runs and updates packages and their dependencies. Conda easily creates, saves, loads and switches between environments on your local computer. It was created for Python programs, but it can package and distribute software for any language. Conda can be combined with continuous integration systems such as Travis CI and AppVeyor to provide frequent, automated testing of your code. The conda package and environment manager is included in all versions of Anaconda and Miniconda. Anaconda Repository. Conda is also included in Anaconda Enterprise, which provides on-site enterprise package and environment management for Python, R, Node.js, Java and other application stacks.

**Applications Available in Navigator**

* + - JupyterLab
    - Jupyter Notebook
    - QTConsole
    - Spyder
    - VSCode
    - GlueViz
    - Orange 3 App
    - Rodeo
    - Rstudio

**Jupyter Notebook**

The Jupyter Notebook is an open-source web application that allows you to create and share documents that contain live code, equations, visualizations and narrative text. Uses include: data cleaning and transformation, numerical simulation, statistical modeling, data visualization, machine learning, and much more.

**Python**

Python is a widely used general-purpose, high level programming language. It was initially designed by Guido van Rossum in 1991 and developed by Python Software Foundation. It was mainly developed for emphasis on code readability, and its syntax allows programmers to express concepts in fewer lines of code. It is a programming language that lets you work quickly and integrate systems more efficiently. There are two major python versions- **python 2 & 3**.

Some of the advantages of using python:

* Emphasis on code readability, shorter codes, ease of writing
* Programmers can express logical concepts in fewer linesof code in comparison to languages such as C++ or Java.
* Python supports multipl**e** programming paradigms, like object-oriented, imperative and functional programming or procedural.
* There exists an inbuilt function for almost all of the frequently used concepts.

**Features**

* Interpreted
* Platform Independent
* Free and open source, Redistributable
* Embeddable
* Robust
* Rich Library support

**Data Analysis in Python**

Python is a great language for doing data analysis, primarily because of the fantastic ecosystem of data-centric Python packages. Pandas one of those packages, and makes importing and analyzing data much easier.

**CHAPTER 5**

**SYSTEM DESIGN**

**5.1 SYSTEM ARCHITECTURE**

Web App

Data Collection

Opinion Extraction

Medium

Response Detection

SVM

Recommendation System

Pre-processing

Data Extraction

No

Yes

Data Base

Data Reduction

Data transformation

Data Cleaning

**Figure: 5.1 System Architecture**

**5.2 MODULE DESCRIPTION**

The proposed system consists of the following modules

* Data collection
* Data preprocessing
* SVM
* Response detection
* Recommendation system

**5.2.1 DATA COLLECTION**

Data is a piece of information that should be collected carefully so that the collected information is useful. Data collection is an important step while doing experiments or researches. The above fig.5.1 illustrates the view of data collection and other phases of the proposed system. Data collection is the process of gathering or processing information that is used for obtaining outcomes in experiments. To select smart village based on collection of different opinion from village people in terms of forms, questionnaire, views and surveys. Opinion mining extracts useful knowledge about village from ample of opinions. This mining process gives us a right direction for creating smart village Detailed data’s or information are collected from the people by the web application. The process done by the uploading the web app in internet. And some aware issued to the village people for our survey. And also, the database is created for the storing purpose. The application contains some questions, that covered some criteria that satisfied which village is smart.

* Criteria based on specification and facilities available in the village.

Cr: 1 Water facility

Cr: 2 Transport facility

Cr: 3 Electricity facility

Cr: 4 Healthcare and nutrition facility

Cr: 5 Government offers

Cr: 6 Clean water availability

Cr: 7 Edu-school/college availability

Cr: 8 100% literate

Cr: 9 Road side plantations

Cr: 10 Individual toilet availability

Cr: 11 Awareness toward organic farming

Cr: 12 Opportunities of employment

Cr: 13 Soil Conservation

Cr: 14 Removal of Environment Pollution

Cr: 15 Facility of Computer Education

Cr: 16 Mobile Communication

Cr: 17 Woman Welfare

Cr: 18 Awareness toward social media

Cr: 19 Social solidarity

Cr: 20 Environment management

The above criteria are the smart village should be satisﬁed among villagers. These factors are required to setup smart village. Initially, collect survey from villagers on above factors. Collected data are stored in the database, then its exported to the CSV format. The dataset covered over some particular region. it means, maximum of the survey done by the PATTUKKOTTAI region people opinion. On the basis of satisfying criteria, to conclude whether village can be ﬁt for smart village or not.

**5.2.2 DATA PREPROCESSING**

Data preprocessing is a data mining technique that involves transforming raw data into an understandable format. Real-world data is often incomplete, inconsistent, to contain many errors. Data preprocessing is a data mining technique that involves transforming raw data into understandable format. Real world data is often incomplete, inconsistent and lacking in certain behaviors or trends, and is likely to contain many errors. Data preprocessing is a proven method of resolving such issues. Data preprocessing is used database-driven applications such as customer relationship management and rule-based applications data preprocessing is a proven method of resolving such issues. Data preprocessing prepares raw data for further processing.

**Data mapping:** In this phase, the text data converted into numerical form. That means the dataset is formatted into some region of values. Here every parameter changed to some region of numerical value.

**Data Normalization:** decomposing tables to eliminate data redundancy(repetition) and undesirable characteristics like Insertion, Update and Deletion Anomalies. It is a multi-step process that puts data into tabular form, removing duplicated data from the relation tables.

**Dataset Splitting:** The structured dataset splitting into x and y sets. x contains factor attributes (our criteria’s), y contains final choice. The final choice is based on satisfying the criteria which village is smart. The Dataset splitting into training and testing dataset. The preprocessed dataset is input for the next level of processing. The dataset splitting into training dataset and testing dataset. It means, dataset spilt into 70% of training set and 30% of testing set. The final choice considered to separate dataset due to our research. The dataset is individually splitting into training and testing dataset.

Text Data

Mapping

Normalization

Splitting training and testing dataset

**Figure: 5.2.2 Data preprocessing**

* + 1. **SVM / MAXIMUM ENTROPY**

**SVM (Support Vector Machine) Classifier**

To build a model able to predict target values of information instances in the testing set, for which only known the parameters. Text data are ideally suited for SVM classification because of the sparse nature of text, in which few features are irrelevant, but they tend to be correlated with one another and generally organized into linearly separable categories is trained to model aspect classification and this trained SVM is used for polarity classification per aspect. The experimental process done by the training and testing dataset. Initially the model was created by the both training dataset and model is called prediction set. The prediction set is compared to the factor testing dataset. Next confusion matrix calcite for the prediction set and final choice testing dataset. Finally, the accuracy is reached to expected range. the results indicate that the proposed techniques have achieved about 100% accuracy. Web based data are applied to emotion cause extraction sub system and complementary feature selection method, based on the output of these features are merged. In training process, web post with unknown emotions are given to SVM and SVR classification model and the output gives information about the type of emotion. SVM models can be classified into four distinct groups:

1. Classification SVM type 1
2. Classification SVM type 2
3. Regression SVM type 1
4. Regression SVM type 2

In Support Vector Machine, initially the model was created by using training dataset. The model is named on prediction dataset. the Prediction process done to the testing dataset. Finally, the confusion matrix applied to the testing dataset and then the accuracy level was check by code. classification accuracy alone can be misleading if you have an unequal number of observations in each class or if you have more than two classes in your dataset.

**5.2.4 RESPONSE DETECTION**

The polarity identification (positive/ negative/ neutral) process predicts by the sentimental analysis techniques. due to the large amount of information and the wide range of sources, it is essentially impossible for a customer to read all of the reviews and make an informed decision on whether to purchase the product. It is also difficult for the manufacturer or seller of a product to accurately monitor customer opinions. For this reason, mining customer reviews, or opinion mining, has become an important issue for research in Web information extraction. One of the important topics in this research area is the identification of opinion polarity. The opinion polarity of a review is usually expressed with values 'positive', 'negative' or 'neutral'. We propose a technique for identifying polarity of reviews by identifying the polarity of the adjectives that appear in them.

**5.2.5 RECOMMENDATION SYSTEM**

At last the final stage, recommendation for selecting smart village depending upon satisfying the factors. This research is achieved by the support vector machine classification. Our goal of the research, specifically to institutionalize a different future for rural villages that reverses the new patterns of rural to urban migration, particularly among young adults, that smart city development has amplified in this region. The classification process is giving the final outcome. After getting output from sentiment analysis techniques, we get different opinions from villagers on different criteria. The villagers are agreed on 20 number of criteria. We can suggest recommending smart village that among 20 criteria; at least 18 criteria should be accepted by villagers. The 18 criteria are major factors for the smart village scheme. Otherwise, we cannot recommend that village for becoming a smart village. Recommendation system, in this part very important phase on our outcome.

**CHAPTER 6**

**CONCLUSION AND FUTURE WORKS**

**6.1 CONCLUSION**

Opinion mining extracts variety opinions from villagers by above methods. To know how much response from villagers in different criteria for smart village, the proposed system uses Support Vector Machine for sentiment analysis of survey data to increase the accuracy. The existing system used Bayes based classifier which assumes that there exists an independency between features. And also, this scheme will be tested using various classifiers such as Decision tree and Neural Networks. Whereas SVM uses a Hyper plane which separates training data into two classes which results in better accuracy in terms of polarity and reach the best accuracy for the performance.

**6.2 FUTURE WORKS**

Classification algorithm relies more on the sentiment analysis of survey data to yield a prediction about research. Future works should consider additional parameters or addition of more criteria based on smart village and other updates in web app to give a better accuracy. Collection of those parameters to the more villages tends to achieve the maximum best result through this research. They should also consider classification is better compared to other.

**APPENDIX 1**

**SAMPLE SOURCE CODE**

**Import file**

>>import pandas as pd

from sklearn.metrics import accuracy\_score

from sklearn.metrics import confusion\_matrix

from sklearn.metrics import classification\_report

from sklearn.model\_selection import train\_test\_split

from sklearn.tree import DecisionTreeClassifier

from sklearn.linear\_model import LogisticRegression

from sklearn.neighbors import KNeighborsClassifier

fromsklearn.svm import SVC

from sklearn.naive\_bayes import GaussianNB

from sklearn.svm import LinearSVC

from sklearn.ensemble import RandomForestClassifier

import matplotlib.pyplot as plt

import numpy as np

**Reading CSV file**

import pandas as pd

df=pd.read\_csv("villasrvey.csv")

df.head()

**Scatter the dataset**

import matplotlib.pyplot as plt

%matplotlib inline

plt.scatter(df['villagename'],df['Education'])

**View Shape of the file**

df. Shape

df.info

**Mapping**

edu\_mapping={'yes':1,'medium':2,'no':0}

df.Education =df.Education.map(edu\_mapping)

df.head()

**Splitting the dataset**

Splitting the dataset into training set and testing set (x\_train , y\_train , x\_test , y\_test)

x=df[['S.No','Education','schoolcollegedifficulties','healthcare','waterfacility','cleanwaterdifficulties','transport','difficultiesinsocialactivities','socialsolidarity','roadfacilities','roadfacilities','electricity','governmentoffer','managementfacilities','employmentdifficulties','awaresocialmedia','library','sanitation']]

y=df['smart']

Enter x, and y

**Training and Testing Data**

from sklearn.model\_selection import train\_test\_split

x\_train, x\_test, y\_train, y\_test =train\_test\_split(x,y,test\_size=0.3)

x\_train

y\_train

x\_test

y\_test

**Scaling**

x\_train[x\_train.dtypes[(x\_train.dtypes=="float64")|(y\_train.dtypes=="int64")].index.values].hist(figsize=[11,11])

**Applying algorithm**

from sklearn.svm import SVC

model\_all=SVC ()

model\_all.fit(x\_train,y\_train)

y\_pred=model\_all.predict(x\_test)

**Confusion matrix**

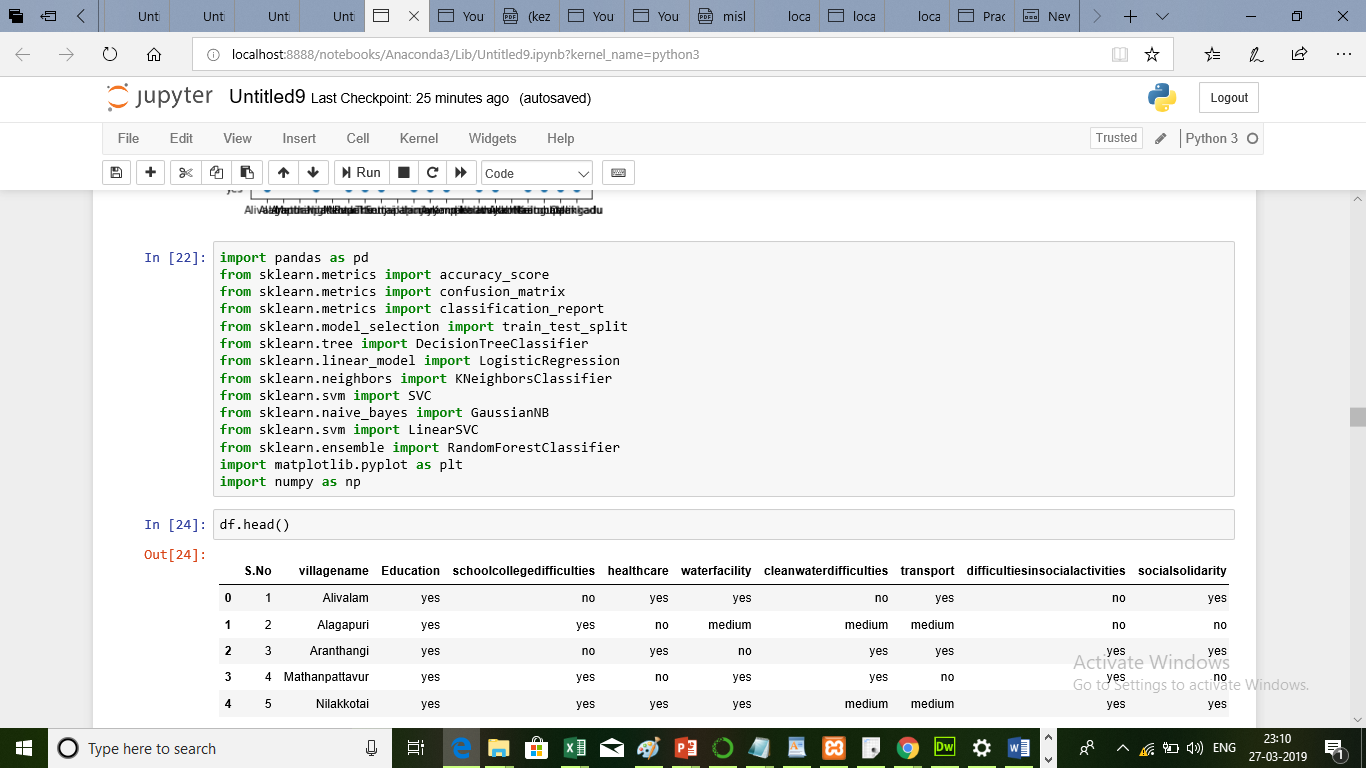
confusion\_matrix(y\_test,y\_pred)

**Accuracy response**

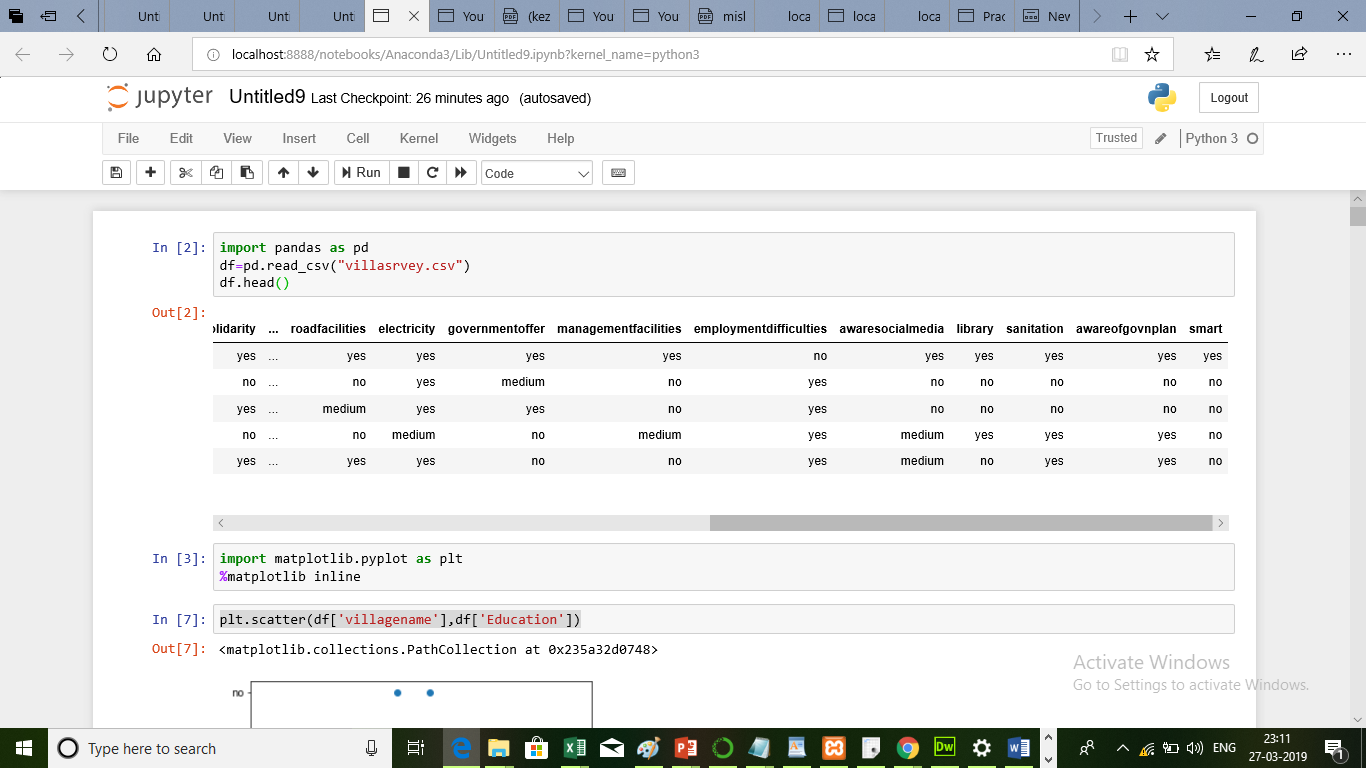
accuracy\_score(y\_test,y\_pred)

**APPENDIX 2**

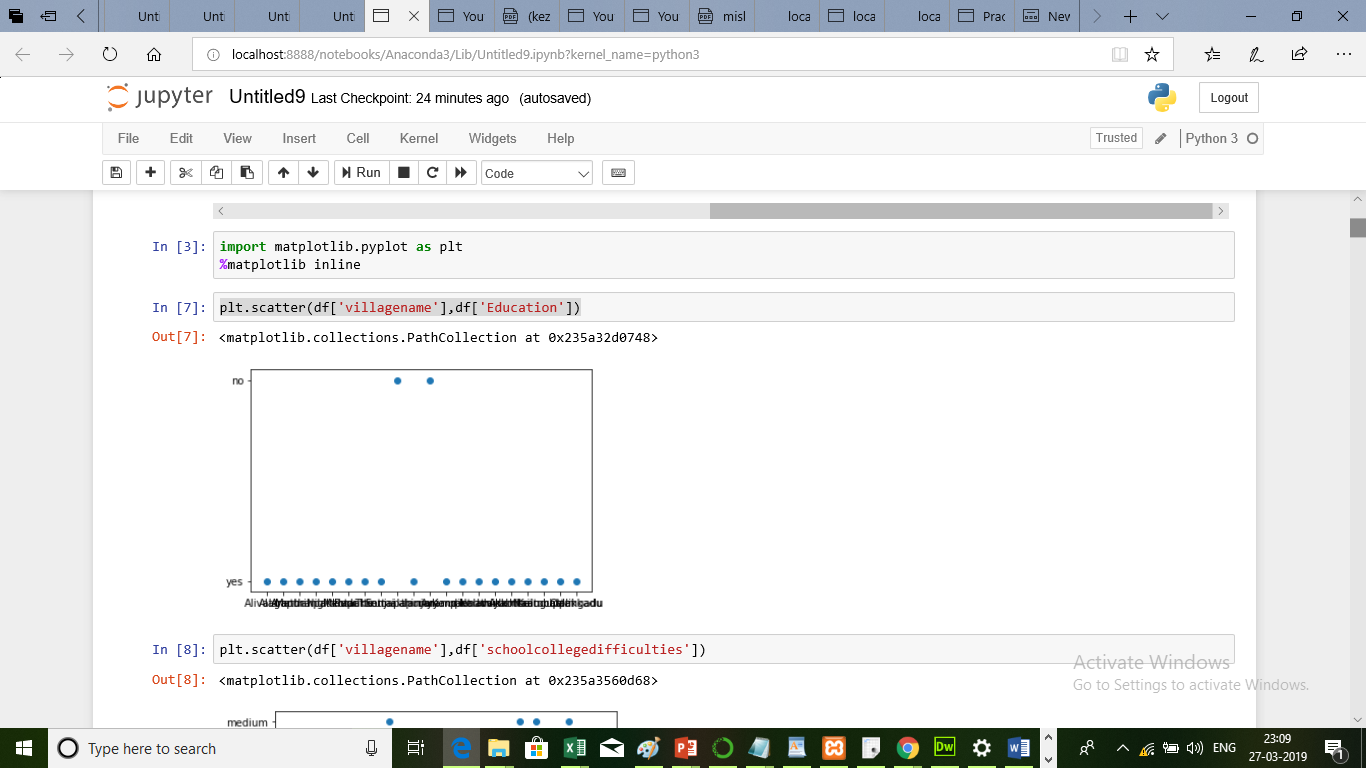
**SAMPLE SCREENSHOTS**

****

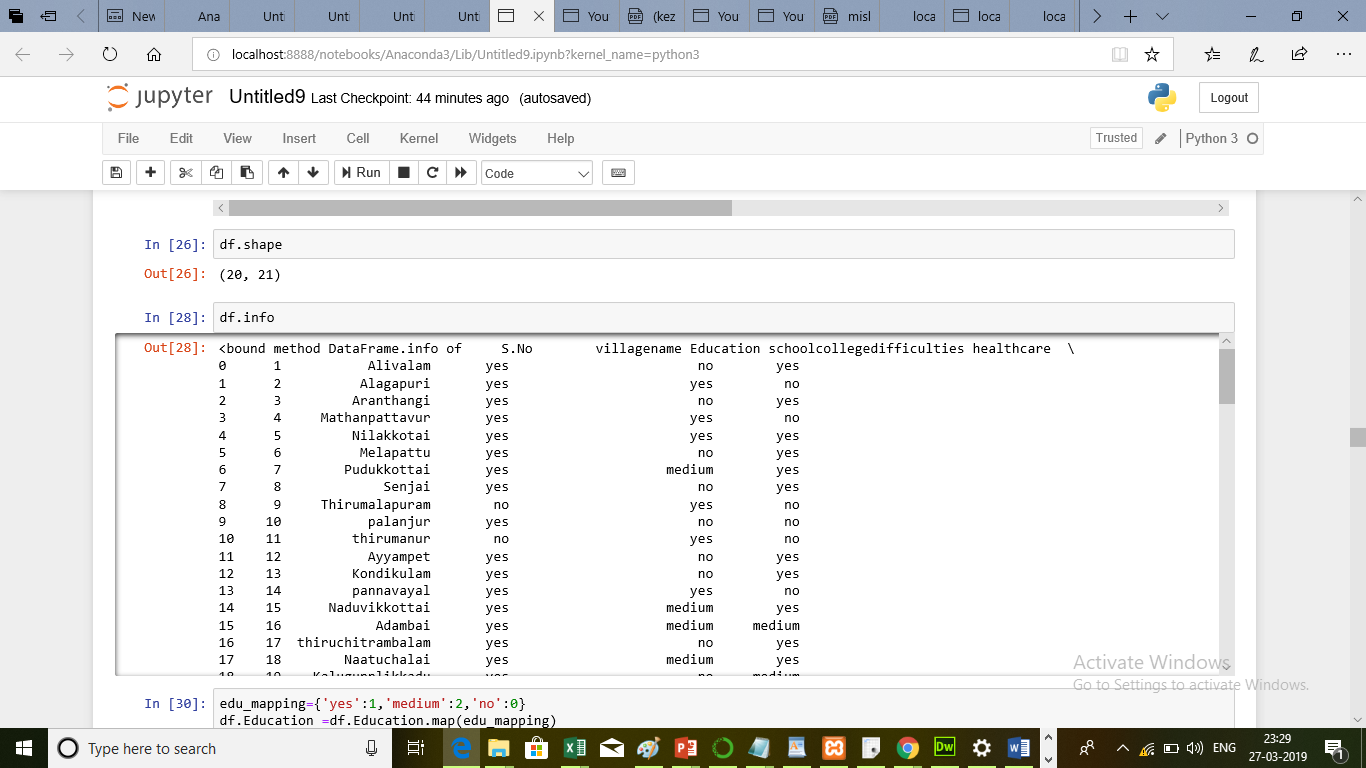
**S-1 : Import file**

****

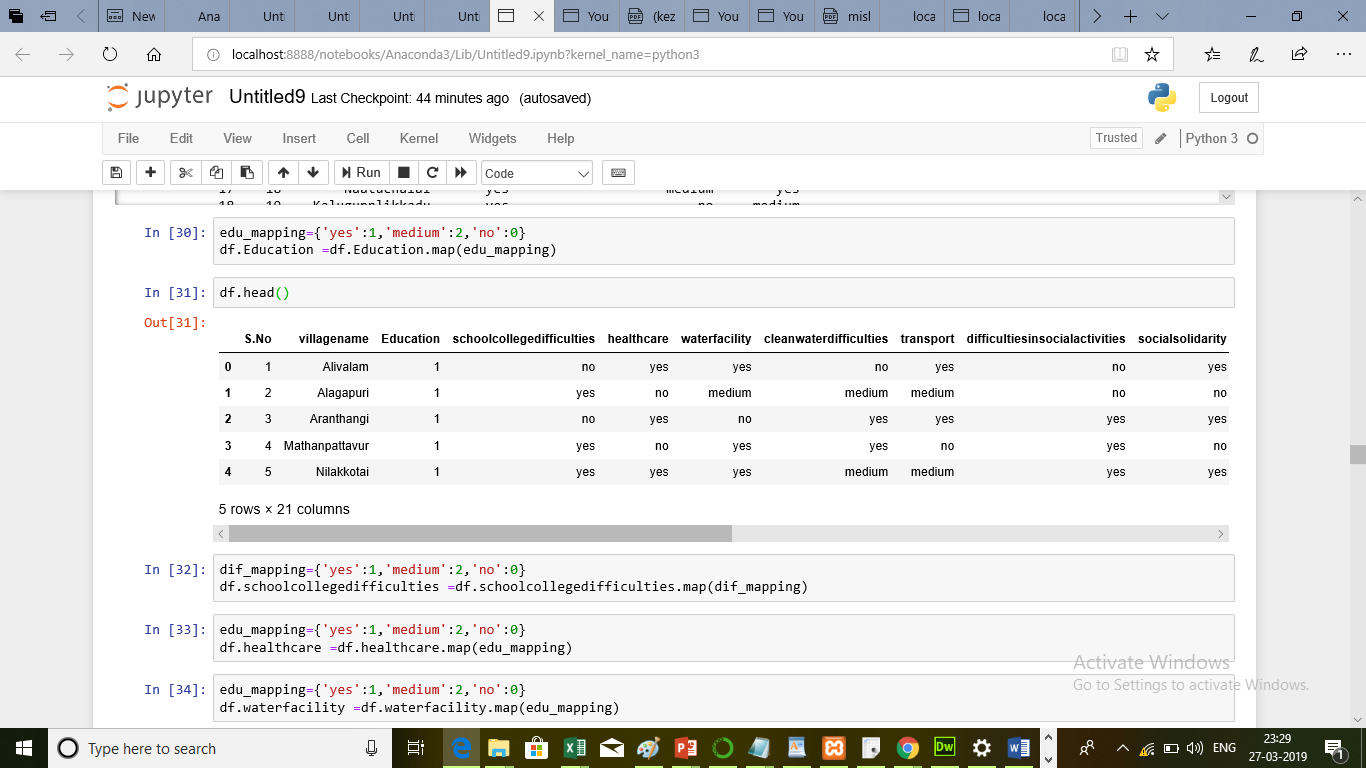
**S-2 : Reading dataset**

****

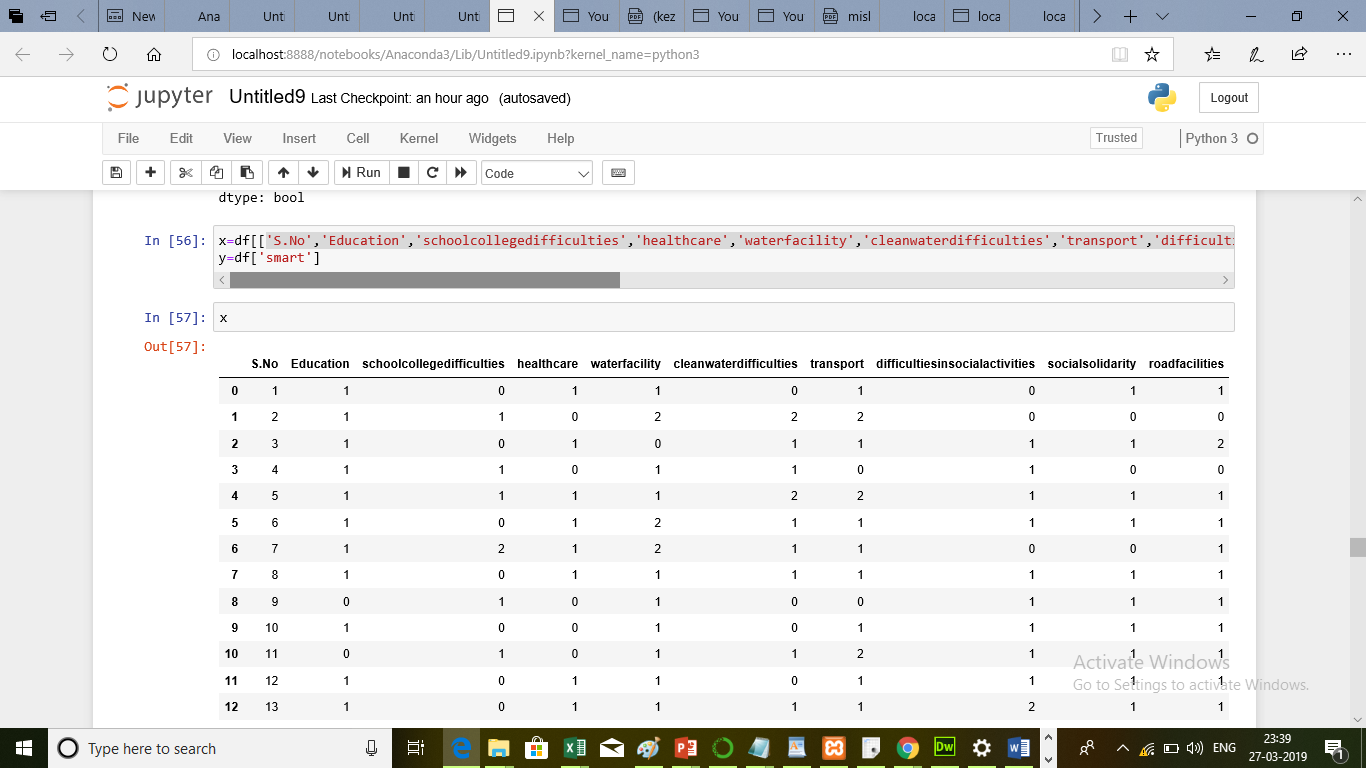
**S-3 : Scatter the attributes**

****

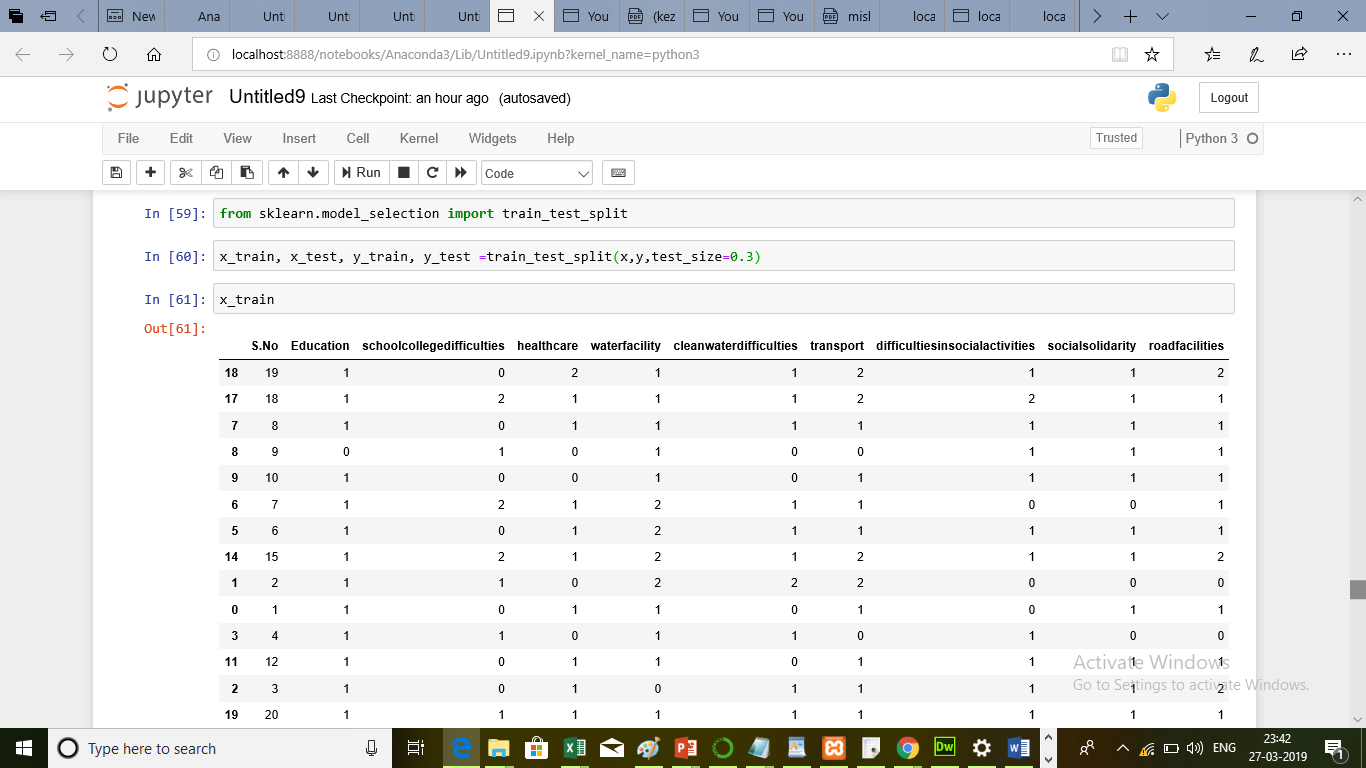
**S-4 :Shape and info of the dataset**

****

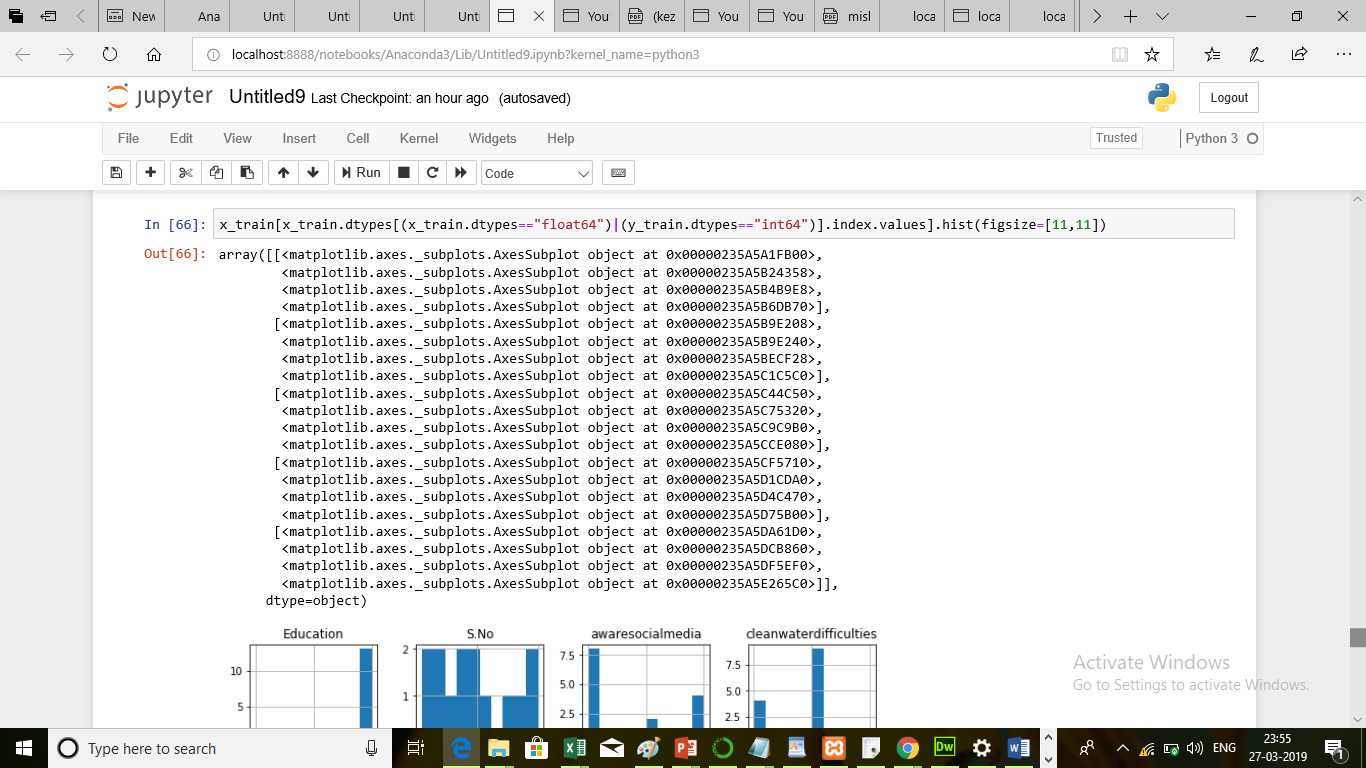
**S-5 : Mapping concept**



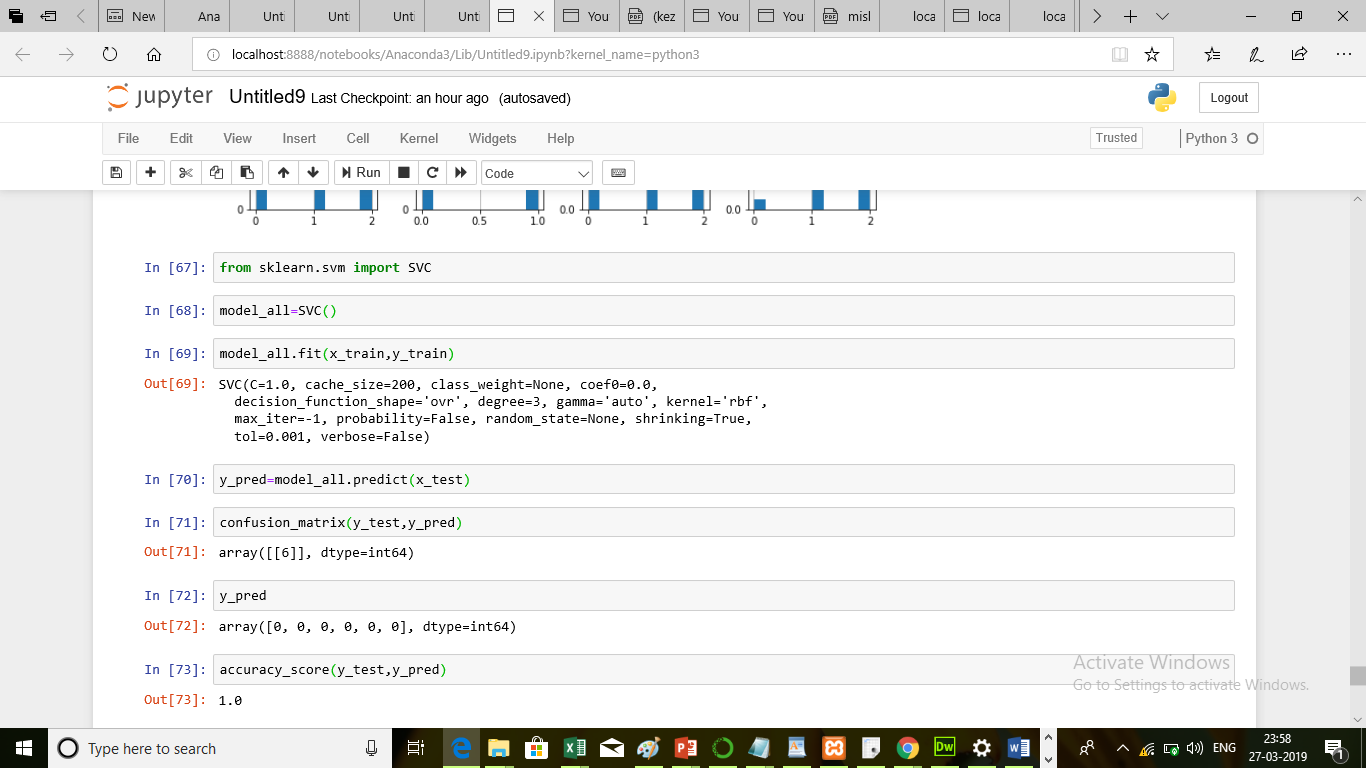
**S-6 : splitting the dataset**



**S-7 : Training and testing dataset**



**S-8 : Scaling**



**S-9 : Applying Algorithm and getting the accuracy.**

**REFERENCES**

1. Kumar. P, Zaidi. A, Shrivastava. A, Jain. G.: Evolving sentiments towards e-governance using opinion mining. Int J Compute Appl (0975–8887), pp. 24–27
2. Sahoo. A.K, Sahoo. K.S, Tiwari. M.: Signature based malware detection for unstructured data in Hadoop. In: 2014 International Conference on Advances in Electronics, Computers and Communications, INSPEC Accession Number: 14841793, pp. 1–6
3. Jaskaran Kaur, Sheveta Vashisht: Analysis and Identifying Variation in Human Emotion Through Data Mining In: Int.J Computer Technology & Applications, 2012
4. A. Jeyapriya, C.S. Kanimozhi Selvi: Extracting Aspects and Mining Opinions in Product Reviews Using Supervised Learning Algorithm IEEE, 2015
5. Hu Minqing, Liu Bing, "Mining and summarizing customer reviews", In Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 168-177, 2004
6. Eirinaki Magdalini, Shamita Pisal, Japinder Singh, "Feature-based opinion mining and ranking", Journal of Computer and System Sciences*,* vol. 78, pp. 1175-1184, 2012
7. Nayak, A.S., Fernandes, R.: A study and analysis of opinion mining and sentiment analysis. Int J Adv Res Compute Eng. Technol (IJARCET). 3(12), pp. 4312–4315 Dec (2014)
8. Vidisha M. Pradhan, et al” A Survey on Sentiment Analysis Algorithms for Opinion Mining” in International Journal of Computer Applications 133(9):7-11. January 2016.
9. Anu Peisker and Soumya Dalai present ed “Data Analytics for Rural Development”. http://www.indjst.org/index.php/indjst/article/download/61494/48039.
10. John DALE, Sunil ISHAIRZAY presented “from Smart Cities to Smart Villages: New Sustainable Futures for Disrupting Rural Migration in MyanmarandIndia” https://isaconf.confex.com/isaconf/forum2016/webprogram/Paper79019.html
11. Ion SMEUREANU, Cristian BUCUR presented by “Applying Supervised Opinion Mining Techniques on Online User Reviews”, https://core.ac.uk/download/pdf/27056535.pdf
12. Akshi Kumar and Abhilasha Sharma “Systematic Literature Review on Opinion Mining of Big Data for Government Intelligence”, Delhi Technological University in Dec 2016.
13. Wala Medhat, Ahmed Hassan, Hoda Korashy, "Sentiment Analysis Algorithms and Applications: A Survey", Ain Shams Engineering Journal, 2014
14. Blessy Selvam1, S. Abirami2, “A Survey on Opinion Mining Framework”, International Journal of Advanced Research in Computer and Communication Engineering, 2013
15. Richa Sharma, Shweta Nigam, et.al “Polarity Detection at Sentence Level” in International Journal of Advanced Research in Computer and Communication Engineering, 2013.