SENTIMENT ANALYSIS OF GARMENTS WORKERS USING MACHINE LEARNING

BY

HASIN WASIN FUAD
ID: 221-25-125

This Report Presented in Partial Fulfillment of the Requirements for the Degree of Master of Science in Computer Science and Engineering

Supervised By

Md Abbas Ali Khan
Assistant Professor
Department of CSE
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

JANUARY 2023

APPROVAL

This Project/Thesis titled **"Sentiment Analysis Of Garment Workers"**, submitted by Hasin Wasin Fuad, ID No: 221-25-125 to the Department of Computer Science and Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of M.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on 17-01-2023.

BOARD OF EXAMINERS

______________________________ **Chairman**

Dr. Sheak Rashed Haider Noori, PhD
Professor and Associate Head
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

______________________________ **Internal Examiner**

Ms. Naznin Sultana
Associate Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

______________________________ **Internal Examiner**

Mr. Md. Sadekur Rahman
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

______________________________ **External Examiner**

Dr. Mohammad Shorif Uddin, PhD
Professor
Department of Computer Science and Engineering
Jahangirnagar University

DECLARATION

I hereby declare that this project has been done by me under the supervision of **Md Abbas Ali Khan, Assistant Professor, Department of Computer Science & Engineering Daffodil International University**. I also declare that neither this project nor any part of this project has been submitted elsewhere for award of any degree or diploma.

Supervised by:

Md Abbas Ali Khan
Assistant Professor
Department of CSE
Daffodil International University

Submitted by:

Hasin Wasin Fuad
ID: -221-25-125
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

First I express our heartiest thanks and gratefulness to almighty Allah for His divine blessing makes it possible for us to complete the final year project/internship successfully.

I am really grateful and wish my profound indebtedness to **Md Abbas Ali Khan**, **Assistant Professor**, Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of my supervisor in the field of "*Machine Learning*" to carry out this project. His endless patience, scholarly guidance ,continual encouragement, constant and energetic supervision, constructive criticism , valuable advice ,reading many inferior drafts and correcting them at all stages have made it possible to complete this project.

I would like to express my heartiest gratitude to **Md Abbas Ali Khan, Assistant Professor, Department of CSE**, for his kind help to finish my project and also to other faculty members and the staff of CSE department of Daffodil International University. I would like to thank my course mates of Daffodil International University, who took part in this discussion while completing the course work. Finally, I must acknowledge with due respect the constant support and patients of my parents.

ABSTRACT

Sentiment Analysis (SA) is a very popular field for researchers in the field of text mining. Computational treatment of opinions, sentiments and subjectivity of texts are shown by sentiment analysis. In this study I examined the sentiment of the most overlooked group of people known as the Garments Workers. First I present an unsophisticated dataset that uses novel reviews of workers to capture their satisfaction. I went to some companies and took data of the Garments Workers manually. Then, I created a document by aggregating worker emotions of the companies and measuring employee sentiment as to whether they are happy or unhappy using the Ensemble Learning approach to count happy and unhappy terms. Finally, I defined worker satisfaction by using a 10 fold validation method. The results of my model suggest that it may be beneficial for investors to incorporate a measure of worker satisfaction into their method for forecasting earnings.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1
INTRODUCTION

1.1 Introduction

With advancements in technology and fields like deep learning, sentiment analysis is becoming more and more common for companies that want to gauge their workers' sentiments. Today, businesses use natural language processing, statistical analysis, and text analysis to identify the sentiment and classify words into positive, negative, and neutral categories. The best companies understand the importance of understanding their workers' sentiments – what they are saying, what they mean and how they are saying. We can use sentiment analysis to identify workers' sentiment in comments, reviews, tweets, or social media platforms where people mention the brand. In this study, I intended a contribution to the literature about applications of text mining within the field of finance. My approach towards garments workers' sentiment analysis was started from the assumption that the garments workers are organizational assets and it is the truth. It is suggested by management studies that industrial culture influences organizational behavior, especially in the areas of efficiency of industries, effectiveness and commitment of workers. From an applications stance, the result may be interesting to garment owners who want to predict garment earnings. Many important accounting researches suggest that kind of information is not properly collected by the garments due to their intangible nature and that hinders the ability to measure the construct itself. For providing evidence to support this, Edmans tracks the "Hundred best Companies to Work for in America" which was published in Fortune magazine. A link between current garment worker satisfaction and future garment earnings which is not immediately visible to the owners can be seen by this. I seek to complement Edmans' work and found evidence to bring it forward that the forecasting power of my model is incremental to the fortune study. Humans are more introverted to express their irritation, dissatisfaction and unhappiness and these help to detect mental satisfaction. By analyzing these sentiments we can easily determine the workers' sentiment that depends on the information of their daily life. If a person expresses his feelings then, methods such as sentiment analysis can be workable. It uses garment workers' data to identify their sentiment. In the end, it can be determined by the users who belong to which categories — Happy or Unhappy through the accuracy of the model. It can have a positive or negative impact.

1.2 Motivation

60–70 million people work in garment factories worldwide. In them, there are 75% garment workers who are female[17]. A majority of those workers are working under informal employment that means their job is not permanent. The ILO has defined this as, "all remunerative work (i.e. both self-employment and wage employment)"[18] that means those which are not registered, protected or regulated by the existing legal or regulatory frameworks, also non-remunerative work which is undertaken in a producer of income enterprise. Those informal workers haven't any kind of secured employment contracts, benefits, representation or social protection. These informal workers are particularly vulnerable to insecurity, especially those who are female workers, as they are female workers, their gender and their precarious working conditions sometimes makes them bound to some places. Addressing violence and discrimination against informal workers is particularly urgent as a matter of scale, since the informal economy remains the main source of employment in the global south. A fast paced globalization of the garment industry has led to significant restructuring, shifting control from manufacturers to retailers who contract the production of their products. It forced garment owners to rely on a controllable labor force and also increasingly cheap to remain competitive, causing sweatshop-like conditions to emerge in factories in the south. The number of women garment homeworkers is increasing day by day in the north also. The conditions of garment workers are generally very poor. In addition, as they are informal workers they feel insecurity and their minimum wage violations are commonly seen, they mostly get unsafe working conditions, they are forced to perform overtime, child labor and pregnancy-based discrimination are mostly common in the garment sectors.

1.3 Rationale of the study

Around the world, so many workers are unhappy in their life and this directly links them to misuse of the tools they use in their daily life. For ensuring that the workers get proper livelihood by which they can live a happy life it is far more important to know the experience of their personal life. It is always hard to identify factors early for designing comprehensive intervention methods to prevent sentiment related issues. As a result, I have proposed and presented a model using a huge dataset to predict the sentiment of the garment workers.

1.4 Research Objective

The real objective of this research is to gain knowledge based on a complex understanding of sentiment in interaction with personality, in order to increase the chance of the workers to have a happy life.

- To offer a happy life for the workers by keeping the records of their sentiment.
- To investigate if they are happy or unhappy in their current situation.
- To create awareness among the industry owners that they can offer some benefits for their workers.
- To introduce a system that can judge and keep the record of workers of a company.
- To save the garment workers from a neglected life.

1.5 Report Layout

The report is divided into six sections. Each chapter addresses multiple aspects of depression risk level. The various sections of each chapter are explained in more detail.

Chapter 1: Introduction

Chapter 1 is one of the most important introductions to this research. In the first chapter Introduction, I will discuss collective intelligence, motivation, the rationale of this study, research questions, and expected output.

Chapter 2: Background & Literature Review

In chapter 2, a complete description of the research background is described. Some related activities were created by the results of this research study, a machine learning system and classification information. A comparative analysis, scope of the problem and identified challenge are also shown in this chapter.

Chapter 3: Research Methodology

In chapter 3, I described all of the similar works and presented the main general strategies that are associated with that work. I did present many algorithmic details that are needed for each implemented method and those were from the mathematical foundations to the current state of affairs.

Chapter 4: Design Specification

In chapter 4, I described the procedure of collecting data, preprocessing of data and determination of elements as well. The principles of rating grouping, design specifications and the results are also detailed in this chapter.

Chapter 5: Implementation

Chapter 5 shows the future scope of this research activity which is briefly outlined in as the scope of this research study. In this chapter I concluded the entire research paper with a useful conclusion and briefly described the main findings of the study.

Chapter 6: Conclusion and Future Scope

I outlined the commitments and reached the conclusion in this part.

CHAPTER 2
BACKGROUND

2.1 Related Works

This study from “Andy Moniz1” and “Franciska de Jong2” which was intended to contribute to the extending literature about the applications connected to text mining within the field of finance. Their approach they possess towards employees' sentiment analysis has begun from the assumption that employees are organizational assets. It is suggested by managemental studies [1] that organizational behavior is influenced by corporate culture, especially corporate efficiency areas, effectiveness and employee commitment. Indeed, according to the former CEO of IBM, "culture is not just one aspect of the game, it is the game" [2]. From an application stance, our results may be of interest to investors seeking to predict firm earnings. Prior accounting research suggests that such information is not properly incorporated by the stock market due to its intangible nature, hindering the ability to measure the construct itself. To provide evidence in support of this Edmans [1] tracks the “100 Best Companies to Work for in America” published in Fortune magazine. The study posits a link between current employee satisfaction and future 520 A. Moniz and F. de Jong firm earnings that is not immediately visible to investors. They seek to complement Edmans’ work and find evidence to suggest that the forecasting power of our model is incremental to the Fortune study. They extended the regression-based approach adopted by [1] to denote the properties of an object that proxies firm outlook. The remaining part of this study is organized as follows: Section 2 provides an overview of the online employee reviews dataset and highlights its advantages over the Fortune dataset. Section 3 defines employee satisfaction by developing the concepts of polarity and aspect. Throughout this paper we use the term sentiment to denote the polarity of employees’ reviews and aspect to denote the properties of an object that are commented on by reviewers. Then they described their approach to determine the classification of employee satisfaction via its impact on future firm earnings. In Section 4 they developed a polarity-only and a joint polarity-aspect model to predict firm earnings. Section 5 provides an empirical evaluation of the proposed model. They concluded in Section 6 and provided suggestions for future research.

They collected employee reviews from the career community website Glassdoor.com. The platform covers more than 250,000 companies around the world that contains almost 3 million unknown salaries and reviews from 2008 onwards [3]. Reviewers produced an overall score on a scale of 1-5 and rated those companies across five dimensions: Career Opportunities, Work/Life Balance, Culture & Values, Comp & Benefits and Senior Management. Some of these ratings only began in 2012. They extracted employees' full reviews, including their perceived pros and cons of the company [4] and their 'Advice to Senior Management'. The opening sentence of reviewers' text follows a format that is properly structured, identifying if the reviewer is a former or current employee together with the years' of service by number. Comments are reviewed by website editors before publicly posted. This prevents reviewers from posting defamatory attacks and from drifting off-topic that may otherwise hinder topic modeling and sentiment analysis [5] [6]. As a means to aid comparability to [1], they restricted their analysis to the companies that are publicly traded and published in Fortune magazine's "100 Best Companies to Work for in America" list. Their corpus comprises 41,227 individual reviews, from them the current employees wrote two thirds of that and the former employees wrote the remainder. The median number of reviews per company is 340, with 84% of company reviews starting in 2008. Unlike the Fortune dataset which suffers both from untimely (annual) updates and limited data coverage, They believe that employee website comments mitigate such issues, provide a richer source of information and a novel way to look inside a company's culture [3]. Their research employs sentiment analysis using a non-proprietary dataset that they make available in open access to encourage further research.

The approach towards employees' sentiment analysis presented here starts from the assumption that employees are organizational assets and they consist of three steps. Firstly, Latent Dirichlet Allocation (LDA) was employed to identify the aspects in the reviews of employees and then manually infer one latent topic that appears to be associated with firm outlook. Secondly, they measured employee sentiment as the polarity of a composite document, defined by aggregating employee reviews for each firm over the fiscal quarters one by one. The General Inquirer dictionary was used by them for counting positive and negative terms. In line with [9], our goal is not to show that a term counting method can perform as well as a Machine Learning method, but to provide a methodology to measure the impact of employee sentiment on firm earnings. Finally they defined satisfaction of employees as a weighted combination of firm outlook and employee sentiment. They developed a regression-based model [8][10] to forecast firm earnings by placing greater weight on documents that emphasize firm outlook.

2.2 Scope of the problem

Multiple machine learning algorithms were used for model training and test datasets. By identifying associations between attributes in my dataset, I tried to identify specific contributors to sentiment factors. The use of automated sentiment analysis is uncommon in our country. Better sentiment analysis algorithms allow the system to detect sentiment of the users, saving time and money and making it easier to track the workers' situation.

2.3 Challenges

Exactly raw data of the feelings of the garment workers were collected in person from Bangladesh from right at the starting of this thesis. The data was collected by me from the datasets from person to person by using a personal dataset sheet. As my dataset is related to personal information of the garment workers I had to work hard for that because everyone was not actually friendly sharing their personal feelings. Many workers hide their privacy but finally I could successfully collect the proper datasets from two of the garments of Bangladesh. In consequence, the data remained incomplete. Correspondingly, this sort of data provides low accuracy.

CHAPTER 3

RESEARCH METHODOLOGY

3.1 Proposed Model

In this section, I discuss the proposed model for forecasting sentiment analysis variables in detail. This section outlines the architecture of the system. The dataset includes personal data of garment workers described in Section 3.2. Section 3.4 describes the machine learning tactics that I studied and employed for the prediction. My research process is depicted in Figure 3.1 below.

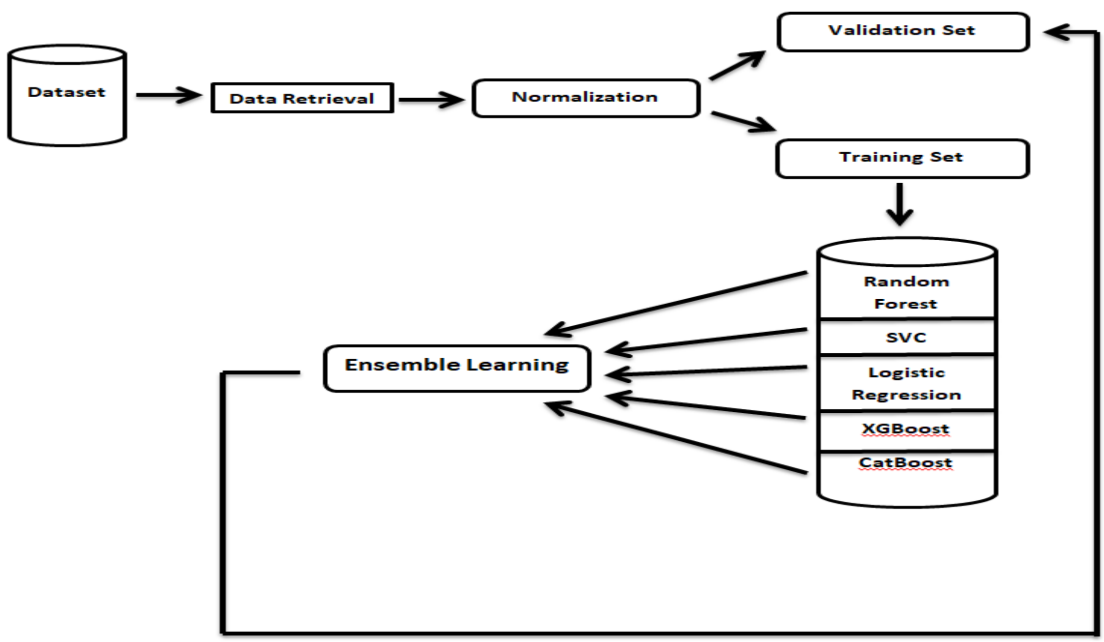

Fig 3.1: Proposed Model

3.2 Data collection

The dataset was collected manually from the garment workers by myself. I have collected my dataset handwritten by visiting two of the garments in Bangladesh for this research purpose. Here,the attributes of the dataset are part of the personal information of the garment workers and their feelings. What is their age, gender, family, overtime, working hour etc. are given in the dataset.

3 .2.1 Dataset

The dataset was collected manually from the garment workers by myself. I've applied a 10-fold cross-validation approach to train and evaluate the model . I also reduced the number of rows in the dataset to make this proper to apply different Machine Learning Algorithms. The figure shown below carries my raw dataset in some rows and columns.

	Gender	Age	Educational Qualification	Salary	Working Hour	Daily Break Hour	Overtime	Shift	Yearly Leave	Marital Status	Label
0	0	22.0	1	10000.0	7.0	1.0	1	1	1	1	0
1	1	21.0	1	9400.0	7.0	1.0	1	1	0	0	1
2	0	25.0	1	9750.0	8.0	1.5	1	0	0	1	1
3	0	26.0	0	10333.0	8.0	1.0	1	0	0	0	0
4	0	27.0	1	10875.0	8.0	1.0	1	1	1	0	1
...	...	...	...	...	...	...	...	...	...	...	...
2527	0	29.0	0	9900.0	8.0	1.0	0	0	1	1	0
2528	0	21.0	1	11836.0	8.0	1.5	0	1	1	1	0
2529	0	32.0	0	9650.0	7.0	1.5	1	0	1	0	0
2530	0	35.0	0	8450.0	7.0	1.0	1	1	1	1	0
2531	0	20.0	0	10698.0	9.0	1.5	1	0	1	1	1

2532 rows × 11 columns

Fig 3.2: Dataset

3.2.2 Data Set Description

The dataset I used questions to assess the information from participants. I have prepared thirteen questions on it. My work was briefly described by me to the participants in the form. What they had to do was also briefly explained by me. In the study, I asked various fundamental questions about the participants, like their salary, marital status , their working hours, and about their age and so on. The structure of the dataset is displayed in the table below.

Table-3.2: Dataset and attributes

Sl. No	Attributes	Type
01	Gender	Nominal
02	Age	Numeric
03	Educational Qualification	Nominal
04	Salary	Numeric
05	Working Hour	Numeric
06	Daily Break Hour	Numeric
07	Overtime	Nominal
08	Shift	Nominal
09	Yearly Leave	Nominal
10	Marital Status	Nominal
11	Label	Nominal

3.2.3 Data Pre-processing

Before developing a prediction model firstly we need to analyze the data. The model's efficiency becomes good if the data is transformed properly. This phase of the model takes care of inappropriate data to obtain more exact and accurate results. The dataset has missing values which cannot be zero values, so it is then normalized to standardize all values. The null values were filled with the mean value of the dataset. Furthermore, some feature values are in the form of strings, which I have converted into numeric types and then standardized the dataset values. After preprocessing the data, I used machine learning classifiers to categorize it. I used two scikit-learn classes to preprocess the dataset

3.2.4 Splitting Dataset

In order to utilize any machine learning technique, the dataset must be divided into two parts: one for model training and the other for model testing. This is referred to as data partitioning. Before employing any machine learning approaches, this must be done. The k-fold cross validation method can be employed for datasets with fewer samples. A data set is broken down into k sets for k-fold cross validation, which makes use of k-1 for training and 1 for cross validation. For my research, I split my data set into 10 folds. In the first iteration, the first fold of the dataset is used as testing data and the remaining folds are utilized as training data. The second fold is used as testing data and the rest of the data is employed as training data in the second iteration. This process is repeated until all 10 folds of all 10 folds have been utilized as testing data. All of the entries in the initial training dataset are used for both training and validating in the k-fold cross validation approach. Additionally, each submission is checked at least once and never more than once.

3.3 Learning Classifiers

I have used various classifiers such as Random Forest Classifier, Logistic regression Classifier, XGBoost,CatBoost, SVC for predicting data analysis.

3.3.1 Random Forest Classifier

To raise the number of more powerful and precise trees for the model, the Random Forest approach employs an immense quantity of trees. To accomplish this objective, ensemble learning with the production of a plurality emitting class can be applied. This figure illustrates the structure of the Random Forest method.

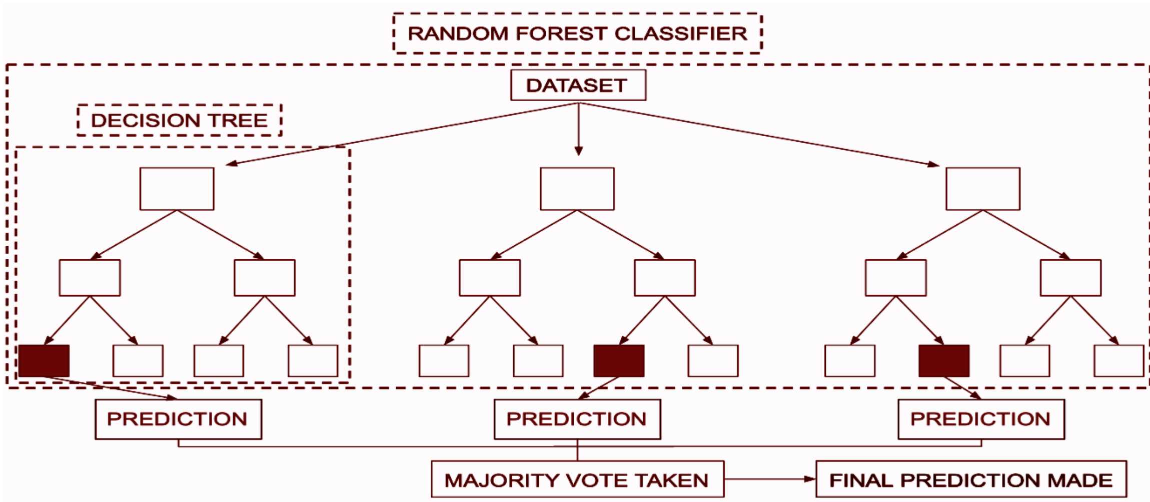

Fig 3.3: Random Forest Classifier

3.3.2 Logistic Regression Classifier

Logistic regression is a type of machine learning classification borrowed from the area of statistics. Logistic Regression is a statistical approach for investigating a dataset where there are one or more independent variables that influence a result. The purpose of utilizing logistic regression is to find the ideal fitting model to explain the relationship between the dependent and the independent variable.

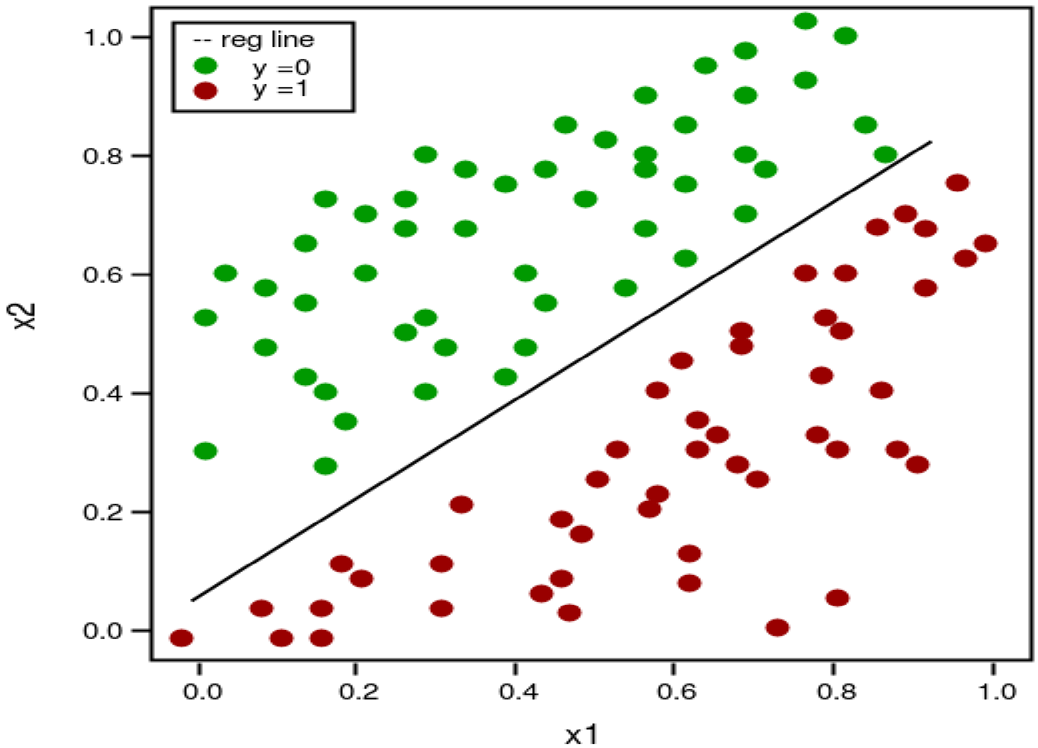

Fig 3.4: Logistic Regression Classifier

3.3.3 SVC Classifier

Support Vector Classifiers (SVCs) are a powerful machine learning approach utilized for both regression and categorization tasks. They are renowned for their indestructibility and capacity to manage convoluted datasets. SVCs are a type of supervised learning that functions by discovering a hyperplane that best separates a set of data into two categories. This hyperplane is identified by maximizing the boundary between the two classes, leading to a model with the most elevated generalization performance. SVCs have been applied effectively in a variety of contexts ranging from text classification to image acknowledgment.

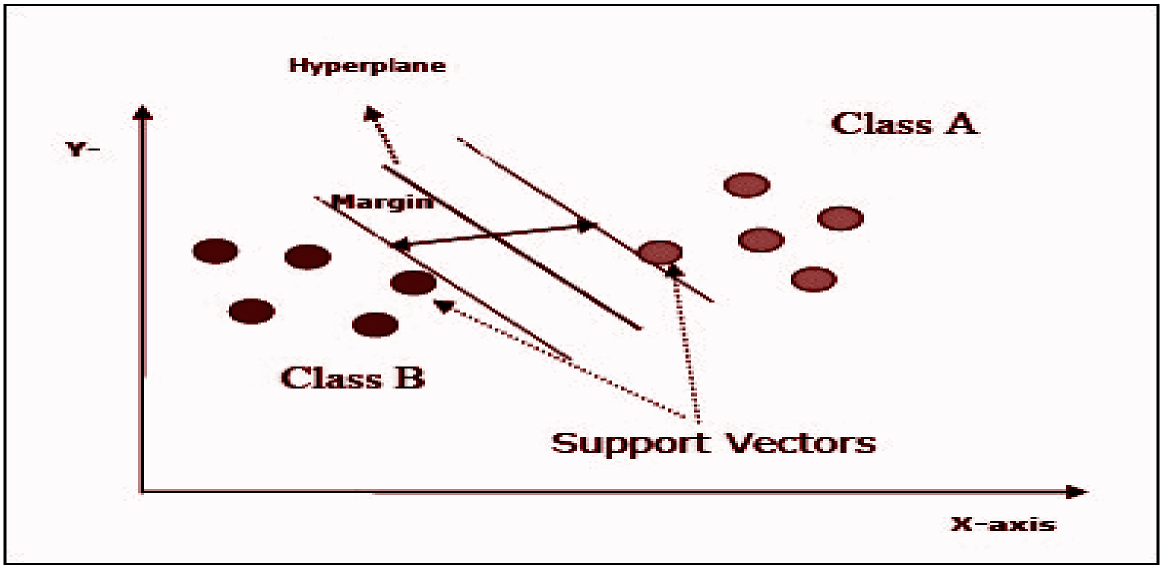

Fig 3.5: SVC Classifier

3.3.4 XGBoost Classifier

XGBoost Classifier is a powerful and reliable machine learning technique that is utilized for both classification and regression tasks. It is an efficient and robust algorithm that is renowned for its precision and velocity. XGBoost Classifier is a gradient-boosting algorithm based on decision trees and can manage a variety of data types. It is an ideal option for large and intricate datasets, as it can take on a large number of features and accurately forecast the target variable. XGBoost Classifier is also cost-efficient, as it does not necessitate a lot of resources. XGBoost Classifier is a great choice for any data scientist or machine learning expert who is seeking an accurate and efficient algorithm for their predictive modeling projects

3.3.5 CatBoost Classifier

CatBoost is a powerful open-source machine learning library designed to deliver quick, precise, and dependable classification. It achieves this by utilizing a combination of decision trees and boosting algorithms to develop strong models capable of accurately forecasting the result of a given issue. CatBoost is highly adaptable, allowing users to optimize their models for different types of data and specific objectives. It also presents a variety of characteristics such as automatic feature selection, model translation, and multi-category classification. CatBoost is an optimal solution for any machine learning project, providing a powerful and dependable classifier for any data driven endeavor.

3.4 Training Model

I trained the model and validated it by using the capabilities of scikit-learn from the Python programming language. Hyperparameter optimization and feature engineering enhanced the performance of the model. A collection of training data was utilized to build the model. Note that feature engineering altered some of the characteristics of the training data set. To begin the training stage, a problem statement must first be constructed, the data set retrieved, and then the data must be cleaned before being sent to the model. Moreover, the algorithms to be employed must be determined, along with the parameters (hyperparameters) they require. After this is completed, the dataset can be divided into a training and testing set, the model algorithms trained, and then tested. My initial training data is a limited resource that must be used prudently. We can use some of it to train the model and other parts to test it; however, the same data can't be used for both. Without giving the model new data that it has never seen before, testing it accurately is not possible. By splitting the training data into two or more sets, I can train and test the model using a single source of data. This allows me to ascertain if the model is overfitting or if it performs well with

3.5 Evaluating Model

The model can be evaluated through a number of metrics, including accuracy, precision, recall, and the F1 score. Accuracy is the proportion of correctly classified documents, while precision is the portion of relevant documents that are accurately classified. Recall is the percentage of pertinent documents that are correctly classified, and the F1 score is the harmonic mean of precision and recall. In addition, the model can be assessed on its ability to correctly classify the sentiment of the research paper, such as positive, negative, and neutral. This can be done by manually labeling a set of documents with the appropriate sentiments and comparing the model's predictions to the ground truth. The model can also be evaluated on its capacity to capture the subtleties of the garment workers research paper, such as the different types of grievances, the intricacy of the topics discussed, and the overall tone of the paper. This can be done by manually annotating a set of documents with the nuances and examining how well the model captures them. Finally, the model can be evaluated on its ability to generate meaningful insights from the research paper. This can be done by manually examining the outputs of the model and evaluating how well it captures the key findings and implications of the paper.

CHAPTER 4

RESULTS & DISCUSSIONS

4.1 Classification Report

The effectiveness of each algorithm was appraised after the Machine Learning Model Creation was applied in order to determine which algorithm was most capable of forecasting depression risk variables. Consequently, the empirical findings form an analytic part in which each possible score for every algorithmic application and approach can be appraised.

Table-4.1: Classification Report

Name Of Classifiers		Percentage				Accuracy
		Precision	Recall	F1 - Score	Support	
Support Vector Classifier	0	0.00%	0.00%	0.00%	0.00	61.4%
	1	1.00%	0.61%	0.76%	254	
	Macro avg	0.50%	0.31%	0.38%	254	
	Weighted avg	1.00%	0.61%	.76%	254	
Logistic Regression Classifier	0	0.08%	0.62%	0.14%	13	62.5%
	1	0.97%	0.63%	0.76%	241	
	Macro avg	0.52%	0.62%	0.45%	254	
	Weighted avg	0.92%	0.63%	0.73%	254	
Random Forest Classifier	0	0.26%	0.53%	0.34%	47	62.5%
	1	0.86%	0.65%	0.74%	207	
	Macro avg	0.56%	0.59%	0.54%	254	
	Weighted avg	0.75%	0.63%	0.67%	254	

XGBoost Classifier	0		0.55%	0.55%	0.55%	99	
	1		0.71%	0.72%	0.71%	155	
	Macro avg		0.63%	0.63%	0.63%	254	
	Weighted avg		0.65%	0.65%	0.65%	254	
CatBoost Classifier	0		0.14%	0.56%	0.23%	25	
	1		0.93%	0.63%	0.75%	229	
	Macro avg		0.54%	0.60%	0.49%	254	
	Weighted avg		0.85%	0.63%	0.70%	254	
Ensembled					97.8%		97.2%

4.2 Predictions

In this study and others, the potency of machine learning models is gauged by prognosticating a variety of iterations. For data analysis prediction, several categorizers, such as Random Forest Classifier, Logistic Regression Classifier, Support Vector Classifier, CatBoost Classifier & XGBoost Classifier can operate independently. I was unable to get the best outcome individually from any of those classifiers and that's why I had to employ the Ensemble Learning technique. The most excellent algorithm for our model can be chosen by comparing them based on accuracy levels attained. From the above Table 4.1 , I can observe that the best algorithm for our model is by uniting two or more classifiers. So, the best result I attained by combining Random Forest Classifier, XGBoost Classifier and CatBoost Classifier.

4.2 Confusion Matrix

Table 4.2: Confusion Matrix of SVC Classifier

Actual	Predicted		
		Positive	Negative
Positive	Positive	0	0
Negative	Negative	98	156

Table 4.3: Confusion Matrix of Logistic Regression Classifier

Actual	Predicted		
		Positive	Negative
Positive	Positive	8	5
Negative	Negative	90	151

Table 4.4: Confusion Matrix of XGB Classifier

Actual	Predicted		
		Positive	Negative
Positive	Positive	54	45
Negative	Negative	44	111

Table 4.5: Confusion Matrix of CatBoost Classifier

Actual	Predicted		
		Positive	Negative
Positive	Positive	14	11
Negative	Negative	84	145

Table 4.6: Confusion Matrix of Random Forest Classifier

Actual	Predicted		
		Positive	Negative
Positive	Positive	25	20
Negative	Negative	73	136

CHAPTER 5
FUTURE SCOPE & CONCLUSION

5.1 Conclusion

Overall, sentiment analysis of garment workers research paper reveals a largely positive sentiment towards garment workers in the areas of wages, working conditions, and job security. While there were some areas of concern, such as pay discrepancies between genders, the majority of research indicates that garment workers have seen improvements in their working conditions in recent years. This is likely due in part to increased government regulations, international standards, and corporate social responsibility initiatives. In conclusion, garment workers have seen significant advancements in their wages, working conditions, and job security in recent years, with more improvements expected in the future.

5.2 Future Scope

Sentiment analysis of garment workers is an emerging area of research that has the potential to improve the quality of life for workers in the garment industry. As technology advances, researchers are able to use sentiment analysis to gain insights into the working conditions and overall satisfaction of garment workers. In the future, sentiment analysis could be used to help identify areas that need improvement, monitor compliance with labor laws, and gain insights into the effectiveness of various labor policies. Additionally, this type of analysis could be used to measure the impact of organizational changes on workers, track the effectiveness of safety and health policies, and understand how workers feel about their work environment. As technology continues to evolve, sentiment analysis of garment workers could be used to develop better labor policies, improve working conditions, and ensure that workers are treated fairly in the workplace.

Reference:

[1] H. Wang, D. Can, A. Kazemzadeh, F. Bar, and S. Narayanan, “A System for Real-time Twitter Sentiment Analysis of 2012 U.S. Presidential Election Cycle,” Proc. 50th Annu. Meet. Assoc. Comput. Linguist., no. July, pp. 115–120, 2012.

[2] P. Goncalves, B. Fabrício, A. Matheus, and C. Meeyoung, “Comparing and Combining Sentiment Analysis Methods Categories and Subject Descriptors,” Proc. first ACM Conf. Online Soc. networks, pp. 27–38, 2013.

[3] B. Pang, L. Lee, and S. Vaithyanathan, “Thumbs up?: sentiment classification using machine learning techniques,” Proc. Conf. Empir. Methods Nat. Lang. Process., pp. 79–86, 2002.

[4] S. J. M. Modha, Jalaj S. , Gayatri S. Pandi, “Automatic Sentiment Analysis for Unstructured Data,” Int. J. Adv. Res. Comput. Sci. Softw. Eng., vol. 3, no. 12, pp. 91–97, 2013.

[5] M. Ahmad, S. Aftab, S. S. Muhammad, and U. Waheed, “Tools and Techniques for Lexicon Driven Sentiment Analysis : A Review,” Int. J. Multidiscip. Sci. Eng., vol. 8, no. 1, pp. 17–23, 2017

[6] Medhat, W.; Hassan, A.; Korashy, H. Sentiment analysis algorithms and applications: A survey. *Ain Shams Eng. J.* **2014**, *5*, 1093–1113.

[7] Razzaq, M.A.; Qamar, A.M.; Bilal, H.S.M. Prediction and analysis of pakistan election 2013 based on sentiment analysis. In Proceedings of the 2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2014), Beijing, China, 17–20 August 2014; pp. 700–703.

[8] Humera Shaziya, G.Kavitha, Raniah Zaheer, 2015, Text Categorization of Movie Reviews for Sentiment Analysis , International Journal of Innovative Research in Science, Engineering and Technology, Vol. 4, Issue11

[9] Akshay Amolik, Niketan Jivane, Mahavir Bhandari, Dr .M. Venkatesan, Twitter Sentiment Analysis of Movie Reviews using Machine Learning Techniques, School of Computer Science and Engineering, VIT University, Vellore

[10] Y. Wu and F. Ren, 2011, Learning Sentimental influence in twitter, Future Computer Science and Application (ICFCSA), 2011, International Conference IEEE vol. 119122.

[11] Edmans, A.: Does the Stock Market Fully Value Intangibles? Employee Satisfaction and Equity Prices. Journal of Financial Economics 101(3) (2011)

[12] Hussaini, M., Kocyigit, A., Tapucu, D., Yanikoglu, B., Saygin, Y.: An aspect-lexicon creation and evaluation tool for sentiment analysis researchers. In: ECMLPKDD (2012)

[13] Kennedy, A., Inkpen, D.: Sentiment Classification of Movie Reviews using Contextual Valence Shifters. Computational Intelligence 22(2), 110–125 (2006)

[14] K. Mouthami, K. N. Devi, and V. M. Bhaskaran, “Sentiment analysis and classification based on textual reviews,” 2013 Int. Conf. Inf. Commun. Embed. Syst., pp. 271–276, 2013.

[15] Kamalapurkar, D., Bagwe, N., Harikrishnan, R., Shahane, S., Gahirwal, M.: Sentiment analysis of product reviews. Int. J. Eng. Sci. Res. Technol. 6(1), 456–460 (2017)

[16]R. Gopal et al,Information mining — reflections on recent advancements and the road ahead in data, text, and media mining,Decision Support Systems(2011)

[17]M. Bansal et al. The power of negative thinking: exploiting label disagreement in the min-cut classification framework

[18]S.M. Kim et al. Automatic detection of opinion bearing words and sentences

[19]L.-W. Ku et al. Major topic detection and its application to opinion summarization
[20] A. Neviarouskaya et al. EmoHeart: conveying emotions in second life based on affect sensing from text

[21]D. Tang, F. Wei, N. Yang, M. Zhou, T. Liu and B. Qin, "Learning Sentiment-specific word embedding for twitter sentiment classification", Proc. 52th Annu. Meeting Assoc. Comput. Linguistics., pp. 1555-1565, 2014.

[22]D. Tang, F. Wei, B. Qin, M. Zhou and T. Liu, "Building large-scale twitter-specific sentiment lexicon: A representation learning approach", Proc. 25th Int. Conf. Comput. Linguistics, pp. 172-182, 2014. Google Scholar

[23]D. Tang, F. Wei, B. Qin, T. Liu and M. Zhou, "Coooolll: A deep learning system for twitter sentiment classification", Proc. 8th Int. Workshop Semantic Eval., pp. 208-212, 2014.

[24]C. D. Manning and H. Schütze, Foundations of Statistical Natural Language Processing, Cambridge, MA, USA:MIT Press, 1999.

[25]D. Jurafsky and H. James, Speech and Language Processing: An Introduction to Natural Language Processing Computational Linguistics and Speech Recognition, Englewood Cliffs, NJ, USA:Prentice-Hall, 2000.

[26]M. Iyyer, J. Boyd-Graber, L. Claudino, R. Socher and H. Daumé, "A neural network for factoid question answering over paragraphs", Proc. Conf. Empirical Methods Natural Lang. Process., pp. 633-644, 2014.

[27]J. Li, R. Li and E. Hovy, "Recursive deep models for discourse parsing", Proc. Conf. Empirical Methods Natural Lang. Process., pp. 2061-2069, 2014.

[28]R. Collobert, J. Weston, L. Bottou, M. Karlen, K. Kavukcuoglu and P. Kuksa, "Natural language processing (almost) from scratch", J. Mach. Learning Res., vol. 12, pp. 2493-2537, 2011.

[29]B. Pang and L. Lee, "Opinion mining and sentiment analysis", Found. Trends Inf. Retrieval, vol. 2, no. 1/2, pp. 1-135, 2008.

[30]B. Liu, "Sentiment analysis and opinion mining", Synthesis Lectures Human Lang. Technol., vol. 5, no. 1, pp. 1-167, 2012.

[31]R. Feldman, "Techniques and applications for sentiment analysis", Commun. ACM, vol. 56, no. 4, pp. 82-89, 2013.

[32]J. Ma, Y. Zhang and J. Zhu, "Tagging the web: Building a robust web tagger with neural network", Proc. 52nd Annu. Meeting Assoc. Comput. Linguistics, pp. 144-154, 2014.

[33]M. Hu and B. Liu, "Mining and summarizing customer reviews", Proc. 10th ACM SIGKDD Conf. Knowl. Discovery Data Mining, pp. 168-177, 2004.

[34]T. Wilson, J. Wiebe and P. Hoffmann, "Recognizing contextual polarity in phrase-level sentiment analysis", Proc. Conf. Empirical Methods Natural Lang. Process., pp. 347-354, 2005.

[35]A. Severyn and A. Moschitti, "Twitter sentiment analysis with deep convolutional neural networks", Proc. Special Interest Group Inf. Retrieval, pp. 959-962, 2015.

[36]
B. Pang and L. Lee, "Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales", Proc. 43rd Annu. Meeting Assoc. Comput. Linguistics, pp. 115-124, 2005.

[37]A. Severyn and A. Moschitti, "On the automatic learning of sentiment lexicons", Proc. Conf. North Amer. Ch. Assoc. Comput. Linguistics, 2015.

[38]R. Collobert and J. Weston, "A unified architecture for natural language processing: Deep neural networks with multitask learning", Proc. 25th Int. Conf. Mach. Learning, pp. 160-167, 2008.

[39]J. Li and D. Jurafsky, "Do multi-sense embeddings improve natural language understanding?", Proc. Conf. Empirical Methods Natural Lang. Process., pp. 1722-1732, 2015.

[40]X. Hu, J. Tang, H. Gao and H. Liu, "Unsupervised sentiment analysis with emotional signals", Proc. Int. World Wide Web Conf., pp. 607-618, 2013.

[41]A. Pak and P. Paroubek, "Twitter as a corpus for sentiment analysis and opinion mining", Proc. 7th Int. Conf. Lang. Resources Evaluation, vol. 2010, 2010.

[42]B. Pang, L. Lee and S. Vaithyanathan, "Thumbs up?: Sentiment classification using machine learning techniques", Proc. Conf. Empirical Methods Natural Lang. Process., pp. 79-86, 2002.
Show in Context CrossRef Google Scholar

SENTIMENT ANALYSIS OF GARMENTS WORKERS USING MACHINE

ORIGINALITY REPORT

20%	%	%	20%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Daffodil International University Student Paper	14%
2	Submitted to Sim University Student Paper	2%
3	Submitted to Higher Education Commission Pakistan Student Paper	1%
4	Submitted to Leiden University Student Paper	1%
5	Submitted to Curtin University of Technology Student Paper	1%
6	Submitted to University of California, Los Angeles Student Paper	1%
7	Submitted to Liverpool John Moores University Student Paper	<1%
8	Submitted to National College of Ireland Student Paper	<1%