**CERTIFICATE**

This is to certify that the Major Project report entitled “**Extraction of verb expressions implying negative opinions**” submitted by

**Name** **Enrollment Number**

Rohit Kumar Singh 00110402713

Deepak Bhandari 00610402713

Shreyas Srivastav 02610402713

is an authentic work carried out by them under my guidance at Amity School of Engineering and Technology, Delhi. The matter embodied in this major project report has not been submitted earlier for the award of any degree or diploma to the best of my knowledge and belief.

------------------------ ----------------------

Prof. M.N. Gupta Mrs. Nidhi Malik

(Head of Department) (Asst. Professor)

**ACKNOWLEDGEMENT**

We are thankful to Prof. Rekha Aggarwal, Director for her moral support and providing us good infrastructure facilities of the institute which helped us in implementing the major project.

It gives us immense pleasure to express our deepest sense of gratitude and sincere thanks to our respected Prof. M.N. Gupta, HOD, Department of CSE & IT and esteemed guide Mrs. Nidhi Malik, Asst. Professor, Department of CSE / IT for their valuable guidance, encouragement and help for completing this work.

We are also thankful to all the staff members of the Department of CSE and IT for their help in our work.

At the end, we would like to express our sincere thanks to all others who helped us directly or indirectly during this project work.

------------------ ----- --------------------- ---------------------

Rohit Kumar Singh Deepak Bhandari Shreyas Srivastav

(00110402713) (00610402713) (02610402713)

**ABSTRACT**

With advanced developments in instrumentation, it has become easier to collect a lot of experimental data in molecular biology. Analysis of such data is very vital as it leads to knowledge discovery that can further be validated by experiments. Cancer is a disease in which abnormal cells divide uncontrollably and destroy body tissues, and every year the number of new cancer patients being registered are over 7 lakh, leaving research in the field of cancer to be our best defense against cancer. Microarray Technology enables a standardized, oncological assessment and prognosis as it generates gene expression profile data that is used for cancer classification and of identification of marker genes associated with cancer. In this work, an application of microarray expression data called classification analysis is employed which uses gene expression data to separate samples into multiple categories. In the work gene expression data is collected, to which filtering and selection is applied by determining the expression ranges of the gene attributes and the data is set then separated into a disparate training and testing set on which the training set is subjected to K-means clustering to identify the set of genes which can be used to determine the probabilistic presence of cancer and test its accuracy with validated data sets.

**Table of Contents**

|  |  |  |
| --- | --- | --- |
| **S.NO.** | **Name** | **Page No.** |
|  |  |  |
|  | **CERTIFICATE** | i |
|  | **ACKNOWLEDGEMENT** | ii |
|  | **ABSTRACT** | iii |
|  | **TABLE OF CONTENTS** | iv |
|  | **LIST OF FIGURES** | vi |
|  |  |  |
| 1. | Introduction | 1 |
|  | 1.1 The Problem of Sentiment Analysis | 4 |
|  | 1.2 Technical Challenges | 7 |
|  |  |  |
| 2. | Literature Review | 9 |
|  |  |  |
| 3. | System Requirement | 12 |
|  | 3.1 Hardware Requirements | 12 |
|  | 3.2 Software Requirements | 12 |
|  |  |  |
| 4. | Text Classification | 13 |
|  | 4.1 Text Mining | 14 |
|  | 4.2 Parts of Speech | 14 |
|  | 4.2.1 Noun | 15 |
|  | 4.2.2 Pronoun | 16 |
|  | 4.2.3 Adjective | 16 |
|  | 4.2.4 Verb | 17 |
|  | 4.2.5 Adverb | 17 |
|  | 4.2.6 Preposition | 18 |
|  | 4.3 Tokenizer | 19 |
| 5. | Extraction of Expressions | 20 |
|  | 5.1 Extraction of Verb Expressions | 21 |
|  | 5.1.1 Verb Extraction Algorithm | 21 |
|  | 5.2 Extraction of Adverb Expressions | 22 |
|  | 5.2.1 Adverb Extraction Algorithm | 22 |
|  |  |  |
| 6. | Training Models | 24 |
|  | 6.1 Naïve Bayes | 24 |
|  | 6.1.1 Representation Used by Naïve Bayes Models | 24 |
|  | 6.1.2 Learn a Naïve Bayes Model from Data | 25 |
|  | 6.1.3 Calculating Probabilities | 25 |
|  | 6.1.4 Make Predictions with a Naïve Bayes Model | 25 |
|  | 6.2 Support Vector Machines | 26 |
|  |  |  |
| 7. | Implementation & Results | 27 |
|  |  |  |
| 8. | Conclusion | 32 |
|  |  |  |
| 9. | References | 33 |

**List of Figures**

|  |  |  |
| --- | --- | --- |
|  | **Figure Name** | **Page No.** |
|  |  |  |
| Fig. 1 | Sentiment analysis process on product reviews | 2 |
| Fig. 2. | Parts of Speech | 15 |
| Fig. 3 | Steps of Implementation | 27 |
| Fig. 4 | Non-negative expression probabilities | 30 |
| Fig. 5 | Negative expression probabilities | 31 |