**PROJECT STATUS REPORT FOR 7th Semester**

1. GROUP NO (If ANY):

CS-05

2. Department/Program:

Computer Science and Engineering

3.

1. Date of Project Report Submission: 17/11/17
2. Extended work of Last (6th) semester: NA
3. New Project/Date of Change (Changed in the 7th semester):Yes

4. MENTOR NAME:

Dr. Rupesh Kumar Dewang

5. Status of the Project (Changes done with respect to your previous reports): NA

6. Project Detail:

1. Title of Last Semester Project/Mini Project: Spam Review Detection Using Okapi Relevance Method
2. Title of New Project: Sentiment Analysis on Hindi Movie Reviews

7. Origin of the Project:

It is human behaviour to look for other's opinion before taking any decision. A lot of documents are available which express opinions on different issues. But the main challenge arises in analyzing these documents to produce useful knowledge. Tremendous works in the area of Sentiment Analysis is available for English language. However, there has been little work done for Indian languages. From the last few years, opinion-rich resources are booming in Hindi and hence there is a need to perform Sentiment Analysis in Hindi.

8. Other Similar ideas available on internet :

* Sentiment Analysis for Hindi Language by Piyush Arora <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.463.7674&rep=rep1&type=pdf>
* Sentiment Analysis in Hindi by Naman Bansal and Umair Z Ahmed <https://cse.iitk.ac.in/users/cs671/2013/submissions/umair/project/report.pdf>
* A Fall-back Strategy for Sentiment Analysis in Hindi: a Case Study by Aditya Joshi et al.

<https://www.cse.iitb.ac.in/~adityaj/HindiSentiWordnet_AdityaJ.pdf>

9. Importance of the proposed project in the context of current status and its relevance to computer science and engineering (Highlight what is the new area or gap which will be solved in the project in relating to what is already known.)

Increasing user-generated content in Hindi on the internet has motivated us to perform sentiment analysis research on movie reviews in Hindi. Sentiment Analysis in Hindi is very challenging because Hindi is a resource scarce language which causes problems in collection and generation of datasets. Also, there are not efficient parsers and taggers for this language. Hindi is morphologically rich and a free order language as compared to English language. Limited resources are available for this language like HindiSentiWordNet (HSWN). It consists of limited numbers of adjectives and adverbs. Thus, this project of ours could prove to be a useful piece of research work in this domain.

10. Work Plan (Prepare a time chart to show Time Schedule of activities)

1. Methodology:

Dataset Collection:

* A dataset of around 250 movie reviews is taken from IIT-Bombay.
* Approximately 750 number of movie reviews are manually collected from movie review site named jagran.com.

Data Preprocessing:

* In this step, all those words which don’t contribute to accuracy such as punctuations, numbers, stop words etc. are removed from the reviews.

Feature Matrix Generation:

We have computed the feature matrix using two methods:

* TF-IDF
* Unigram model

Classification:

We have used following classifiers to predict the polarity of each movie review:

* Neural Network
* Deep Belief Network
* Decision Tree
* Support Vector Machine

1. Time Schedule of activities:

Topic research and study: 1 month

Dataset collection: 2 weeks

TFIDF implementation: 1 week

Unigram implementation: 1 week

Implementing all the classifiers: 1 week

Documentation: Throughout the semester as per progress

1. Outcome expected from the project and its relevance to computer science and engineering.

The project is expected to efficiently classify the movie reviews as positive or negative. The prediction is done on movie reviews in Hindi which is one of the booming language in Web. Much work has not been done in this domain. This project could prove to be different from commonly applied sentiment analysis on English.

1. Summary of roles/responsibilities of all students:

**• Saloni Juneja**

The main role of the student is to implement resource based classification model using Hindi SentiWordNet , implementing various classifiers and comparing accuracies obtained through them. . She was also involved in documentation.

**• Shubham Kumar Goyal**

The main role of the student is to implement feature matrix generation using TF-IDF and unigram models. He also helped in implementation of classifiers like neural network and deep belief network.

**• Sonali Agarwal**

The main role of the student is in dataset collection. She prepared project-status report and project report along with group leader.

**• Saransh Agarwal**

The student helped in implementation of feature matrix generation. He also tried implementing classification of Hindi reviews by translating them into English first.

**• Rohit Kumar**

The student helped in making project report, dataset collection and comparing accuracies and plotting graphs.

Sr.No Regno Name of student Any observation

1. 20144057 Saloni Juneja
2. 20145068 Shubham Kr. Goyal
3. 20144017 Sonali Agrawal
4. 20144119 Saransh Agarwal
5. 20144110 Rohit Kumar

Comments (if any):

Suggestions for improvement (if any): \_ \_\_\_\_\_\_\_\_

Signature of Mentor

PANEL COMMENTS

Comments (if any):

Suggestions for improvement (if any): \_\_\_\_\_\_\_\_\_\_\_

Signature of Panel Representative