# CSE 881 FINAL PROJECT (Cover Page)

Project Title: A Model for Determining Tweet Popularity via Prediction Methods

Project Type (choose one):

Empirical Study/Algorithm Development/Application/Prototype Development

Difficulty Level (choose one): Simple / Moderate / Hard

Justification for your rating: The final project description for the class lists Application type projects as Simple to Moderate difficulty levels, as they do not involved the difficulty of the front end involved in the Prototype Development projects. For this project, the students did not have the opportunity to implement the stretch goal of feature generation from the raw Twitter API features, thus making it a Simple project. However, significant time was spent in the preprocessing stage (both scripting and manual) to get all Twitter features in the correct format for data mining.

Summary of Team Member Participation:

Fill out the following table for each team member with a rating from 1 to 3 (1: poor, 2: satisfactory, 3: good). For “responsive to emails” and “attendance at project meetings”, the rating must be provided by averaging the rating provided by other members of the group.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Name** | **Responsive to emails\*** | **Attended project meetings\*** | **Participate in data collection /preprocessing** | **Participate in coding** | **Participate in analysis/ experiments** | **Writing final report** | **Class presentation** | **Completed Assigned Tasks** |
| Farhan Hormasji | 3 | 3 | 2 | 3 | 3 | 3 | 3 | 3 |
| Bonnie Reiff | 3 | 3 | 3 | 3 | 2 | 3 | 3 | 3 |

Team Member Roles and Contributions:

|  |  |
| --- | --- |
| **Name** | **Roles and Contributions** |
| Farhan Hormasji | Responsible for leading predictive analysis activities as well as contributions to the final report and the project presentation |
| Bonnie Reiff | Responsible for data collection and pre-processing as well as contributions to the final report and the project presentation |

I approve the content of the final report (please add your signature below):

Farhan Hormasji: --------------------------------------------------

Bonnie Reiff : --------------------------------------------------

**A MODEL FOR DETERMINING TWEET POPULARITY   
VIA PREDICTION METHODS**

Farhan Hormasji

Dept. of Computer Science  
Michigan State University

hormasji@cse.medu.esu

Bonnie Reiff

Dept. of Computer Science  
Michigan State University

reiffbon@cse.msu.edu

**ABSTRACT**

A good abstract should include the following information. First, what is it that you are trying to accomplish in this project? Second, why is it useful or important? Third, summarize the results you have obtained.

**Keywords**

Twitter, SVM, ...

# INTRODUCTION

Use this document as a template for writing your final project report. Feel free to modify the organization of the paper where necessary or ignore them if it is not applicable to your project.

For introduction, you should try to include the following:

1. Start by providing a background description of the problem. For example, explain the application domain and discuss how data mining can help better understand the domain. Explain briefly the method you use (e.g., we apply algorithm Y to detect X in this project). You should also summarize the findings of the project (e.g., we demonstrate that X can be detected with accuracy around 85%).
2. **Figure 1**: Taxonomy of classification techniques

   Summarize the overall challenges and technical contributions of the paper in bullets. For example, you may say: The main contributions of this paper are as follows: (1)… (2)…, etc. The more contribution you can think of, the better it is.
3. Provide a road map to the remainder of the paper. Example: Section 2 provides a formal statement of the problem and related work. Section 3….

The total length of the paper should be no less than four pages (double-column format).

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

*CSE881-2015*, Month 1–2, 2004, City, State, Country.

Copyright 2004 ACM 1-58113-000-0/00/0004…$5.00.

# PROBLEM STATEMENT

This section is devoted to providing a formal statement of the problem. If your project is application-oriented, describe the problem domain, the type of data you have collected, and the type of data mining analysis you want to perform. Give a formal definition of the data mining task to be performed (e.g., what is classification, clustering, anomaly detection, association analysis, etc.). If your project also develops a new technique or extends a previous technique, give a brief overview of the previous technique.

## Related Work

You are expected to do a reasonably thorough literature search on the topic you are working on. If your project is application-oriented, give references related to the application domain, references related to previous analysis performed on the domain (if there is any), and references related to the data mining techniques you are planning to use.

If your project is on algorithm development, please make sure that you provide enough references on (1) pioneer algorithms for solving the particular data mining task (e.g., Apriori algorithm for association analysis, statistics-based outlier algorithms for anomaly detection, C4.5 paper for decision tree classification, etc.) (2) State-of-the-art algorithms developed for solving the data mining task.

A good survey of related work should try to organize the references into a taxonomy structure. For example, if you’re doing classification, how would you organize the various classification methods? An example is shown in Figure 1. A good related work must not only list the references, but also describe the pros and cons of the previous work.

# METHODOLOGY

This section should present details of the methodology you have implemented for the project. If a new technique is proposed, present the details of the new technique. If the project is an empirical study, describe all the techniques you plan to investigate in the study. If it is an application or prototype development study, describe the preprocessing and data mining steps you have performed to carry out the project. For prototype development, list all the functionalities of the website that have been implemented.

# EXPERIMENTAL EVALUATION

This section describes the experimental setup and results you obtain.

## Experimental Setup

This section should include:

1. Characteristics of the data set (before and after preprocessing). How many data instances? How many attributes?
2. Evaluation measures you have used
3. Computing platform (operating system and hardware)
4. Software used to generate the results

## Experimental Results

This section should include:

1. The URL of your web site (if you do prototype development).
2. The experiments you had conducted and the results you had obtained

## Discussion

This section should discuss about the significance of the results or any new unexpected insights revealed by the results.

# CONCLUSIONS

Summarize the overall contributions and provide suggestions for future work.

# REFERENCES (at least 5 references)

1. Bowman, B., Debray, S. K., and Peterson, L. L. Reasoning about naming systems. *ACM Trans. Program. Lang. Syst., 15,* 5 (Nov. 1993), 795-825.
2. Ding, W., and Marchionini, G. *A Study on Video Browsing Strategies.* Technical Report UMIACS-TR-97-40, University of Maryland, College Park, MD, 1997.
3. Fröhlich, B. and Plate, J. The cubic mouse: a new device for three-dimensional iput. In *Proceedings of the SIGCHI conference on Human factors in computing systems   
   (CHI ’00)* (The Hague, The Netherlands, April 1-6, 2000). ACM Press, New York, NY, 2000, 526-531.
4. Lamport, L. *LaTeX User’s Guide and Document Reference Manual.* Addison-Wesley, Reading, MA, 1986.
5. Sannella, M. J. *Constraint Satisfaction and Debugging for Interactive User Interfaces.* Ph.D. Thesis, University of Washington, Seattle, WA, 1994.

Columns on Last Page Should Be Made As Close As Possible to Equal Length