# KweriME
# A Q&A system based on topical reputation

Kaushik N
Vasundhara Singh
Shalini Shekar
Harini

Prof. EvLin VidyuLatha P
Batch No. - 16

November 8, 2017

# Problem Statement / Definition

- The dissertation aims to address the issues that reside in the community based Q&A websites with KweriME, a reputation based QA system which employs a category and theme based reputation management system to evaluate users willingness and capability to answer various kinds of questions, while at the same time improving the response latency and answer quality.
- A finely made distributed QA system should facilitate a hike in efficacy of parameters such as question response rate and answer quality, while also trying to deliver a spam-free environment for users. The reputation systems employed by previous works evaluate a user with an overall rating for all questions the user has answered regardless of the question categories, thus it does not accurately reflect the users ability to answer a question in a specific category.Moreover, due to continuous growing size of contents on QA sites, forwarding questions to experts who are willing and able to provide satisfying answers is crucial in maintaining the performance of QA systems.

# Motivation of the Work

Motivation for our work includes the following:

- Seeking information on the internet has become a daily part of our lives. This inspired us to build an efficient system which reduces delay as well as gives us additional information on the content.Most of the Q&A systems experience a delay which we try to solve by forwarding the questions to experts thus, ensuring improved answer quality and also incorporate an anti-spammer control to filter out irrelevant and advertising verbiage to optimize user interaction.Instead of going through multiple questions, the asker will be satisfied with the best and optimal answer, thus, saving time and energy of the asker.
- Previous works include the asker to review the answers manually, thus consuming a lot of time as the answers being posted on these sites grow rapidly. This becomes a tedious job which can be solved by an efficient system. Exploring the talent & knowledge of the answerer by labeling them as experts & providing them a good platform to showcase their talent.

# Literature Survey

Topical interest and Recommending the best answer are the talk of the town which has attracted many researchers interest.

- In [1], the activeness of users has been explored in CQA.They have shown how badges and reputation scores are related to the activeness in different forums based on statistical analysis.Yuhua Lin et al. discussed on clustering the users of Stack Overflow into four clusters namely naive,surpassing, experts and out shiners based on characteristics accounting various metrics by using machine learning algorithm in order to predict the users activities.
- In [2], Tirath Prasad Sahu et al. has worked on the goal of uncovering topic interest, main discussion topics and technology trends over time with the help of statistical topic modeling and also they have worked on the goal of uncovering topic interest, main discussion topics and technology trends over time with the help of statistical topic modeling. In [2], the Questions posted on Stack Overflow has been analyzed both quantitatively and qualitatively in order to improve the success of CQA.

# Methodology

The approach towards the problem definition is divided into following major components:

- **Data Selection :**

  We perform an extensive empirical analysis on a QA based online community dataset to answer the three research questions. Our findings will suggest that:
  i) Prior involvement of the answerer on question tags and topics increases the chance to give the answer for that question.
  ii) Expertise will increase the chance in acceptance of the answer.
  iii) Topical compatibility between the question and answer increases the satisfaction of asker or community with that answer.
  Furthermore, we use various statistical methods in order to implement the algorithms to predict acceptability of the answer by the asker or community.

# Methodology

For instance, we consider the placeholder for the aforementioned online community to be StackOverflow, which is an interactive CQA site for exchanging the knowledge in the software engineering field. It provides a wide variety of functionality for users to gain knowledge in their respective domains. In StackOverflow, there are many questions from various topics related to programming. There are about 10M questions, 17M answers, 4.5M users and 42K tags in the StackOverflow till May 31, 2015. StackOverflow offers its data publicly which is available through Stack Exchange Data Explorer and XML format data dump under creative common licence. The statistics about the dataset relevant to our study are to be analyzed and studied in order to make further inferences on prediction accuracy.

# Methodology

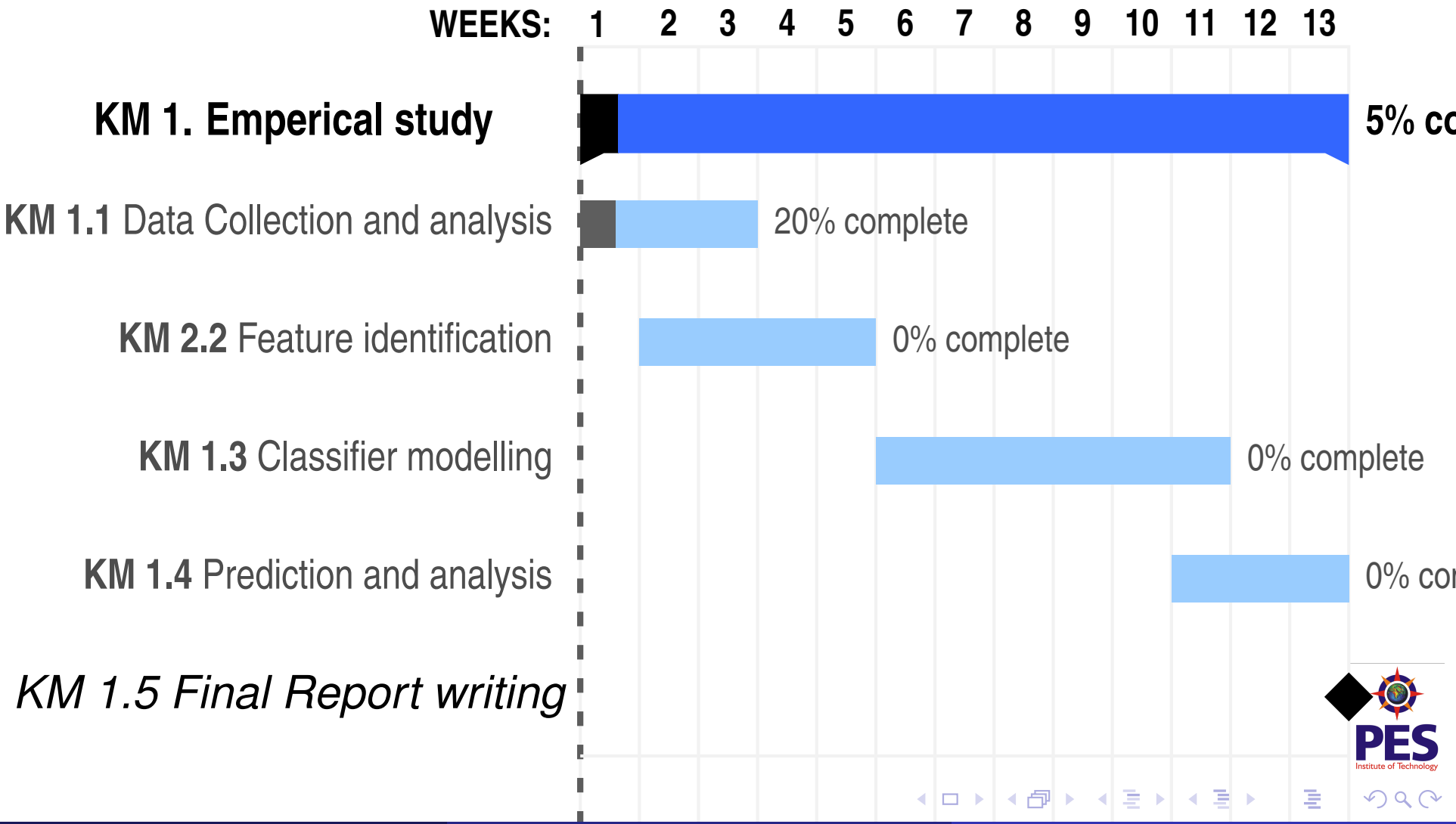
- **Applying Machine Learning techniques :**

  Numerous number of Machine Learning techniques are available which can be used for processing of data such as Least Discriminant Analysis(LDA) which separates different classes by creating a linear boundary,Support Vector Machines(SVM) which uses optimal hyper lane for binary classification, Principle Component Analysis(PCA) is used in case of real, continuous and normalized data, Decision trees which uses a tree structure for classification as well as regression purposes etc.

- **Prediction :**

  Using Machine Learning we predict efficient answers for any query.

# Time line of completion of project from Nov 2017-April 10 2018(Gantt Charts).



# Expected Outcome/ Results

We study and analyze the answers with their questions to predict whether the answer will get accepted or not. We perform an extensive empirical analysis on the retrieved dataset to identify and extract features and perform correlations between various variables to know the causality. Our findings will suggest that:

- Prior involvement of the answerer on question tags and topics increases the chance to give the answer for that question.
- Expertise will increase the chance in acceptance of the answer.
- Topical compatibility between the question and answer increases the satisfaction of asker or community with that answer.

## Outcome

Armed with this observation, we next use classification algorithms to predict acceptability of the answer by the asker or community.So,the outcome of this study lies around predicting the acceptance of the answer as the best answer and various other performance metrics and accuracy ratios that deal with it.

# References


Yuhua Lin, Haiying Shen (2015)

SmartQ: A Question and Answer System for Supplying High-Quality and Trustworthy Answers

*Journal of Latex Class Files* Vol. 14, NO. 8, August 2015.

Tirath Prasad Sahu, Naresh Kumar Nagwani, Shrish Verma (2016)

Selecting Best Answer: An Empirical Analysis on Community Question Answering Sites

*Date of publication : August 26 , 2016*
Digital Object Identifier 10.1109/ACCESS.2016.2600622

# The End