**Project Proposal**

**Team: EXPLORERS**

(at most two-pages, 11-point font, single spacing, 1 inch margins)

1. **Title of the Project: 1. Analyze & Predict Food Review sentiment**
2. **Brief on the project**

The purpose of the analysis is to understand various emotions /sentiments displayed by the customers while providing their reviews on Food .Along with this also understand triggers for the sentiments. By understanding the triggers, companies can retrospect their service/product offered w.r.t to the reviews and make necessary changes impacting customer experience thus impacting business. Also it has been noted that many customers fail to provide ratings while leaving feedback or provide incorrect ratings which impacts the overall score. Hence it is necessary to predict the type of rating and provide a prejudice-free score simply based on comments.

1. **Deliverables of the project:**

The 2 primary Deliverables of the project will include

1. Sentiment analysis against the reviews
2. Predict ratings/sentiments of the reviews

There may also be secondary deliverables depending on the observations made during the analysis

1. Identify triggers for sentiment analysis- product /service
2. Any other observations

In order to analyse data some Pre-processing steps will include: Tokenize, Split into sentences, POS Tag

1. **Resources**
   * **Data set source:** 
     + Kaggle:
       - <https://www.kaggle.com/snap/amazon-fine-food-reviews>
       - The Food Reviews dataset consists of approx 500,000 food reviews users left up to October 2012
   * **Soft ware**: Software you will choose to solve the problem.
     + Python /R programming language
     + SQLite
     + Visualization Tools (ShinyR )
   * **References**:
     + [**https://journalofbigdata.springeropen.com/articles/10.1186/s40537-015-0015-2**](https://journalofbigdata.springeropen.com/articles/10.1186/s40537-015-0015-2)
     + [**http://fjavieralba.com/basic-sentiment-analysis-with-python.html**](http://fjavieralba.com/basic-sentiment-analysis-with-python.html)
     + Kaggle
2. **Team Members:** Names, E-mail Ids and Mob of the people working on the project. Maximum team size is four participants.

|  |  |  |
| --- | --- | --- |
| Names | Emails | Mob |
| Srinivasa Vadlamani | **vadlamani64@yahoo.com** | **+91 9701274068** |
| Suraj Shukla | **surajshukla1105@gmail.com** | **+91 9561024550** |
| Prakash Shetty | **prakash\_2402@yahoo.co.in** | **+91 9819638958** |
| Merlyn Kulkarni | **kmerlyn@gmail.com** | **+91 9819299884** |

1. **Milestones with timeline**

|  |  |  |  |
| --- | --- | --- | --- |
| **Milestones** | **Start date** | **Finish date** | **Grading Grading** |
| 1. **Define a problem** | **15­Sep­16** | **18­Sep­16** | **5%** |
| 1. **Get the Data** | **15­Sep­16** | **26­Sep­16** | **5%** |
| 1. **Explore and pre-process data** | **27­Sep­16** | **23­Oct­16** | **10%** |
| 1. **Create Features** | **24­Oct­16** | **20-Nov­16** | **5%** |
| 1. **Create Model** | **31­Oct­16** | **30­Nov­16** | **15%** |
| 1. **Deploy & consume model** | **21­Nov­16** | **11­Dec­16** | **10%** |
| 1. **Report Writing** | **05­Dec­16** | **15­Dec­16** | **10%** |
| 1. **Project submission** | **15­Dec­16** | **15­Dec­16** | **15%** |
| 1. **Final presentation with project/ product demonstration** | **16­Dec­16** | **18­Dec­16** | **20%** |
| 1. **Blog publishing** | **19­Dec­16** | **23­Dec­16** | **5%** |

**\*Refer Data Science Process Doc – if deadline is missed no weightage will be given and it will entail loss of grade for the group. All milestones should be submitted online as per date indicated by the group.**

1. **Technical architecture**

Defi

Define Problem

Pre-Process data

Create Model

Test Model

Visualization

Define Problem