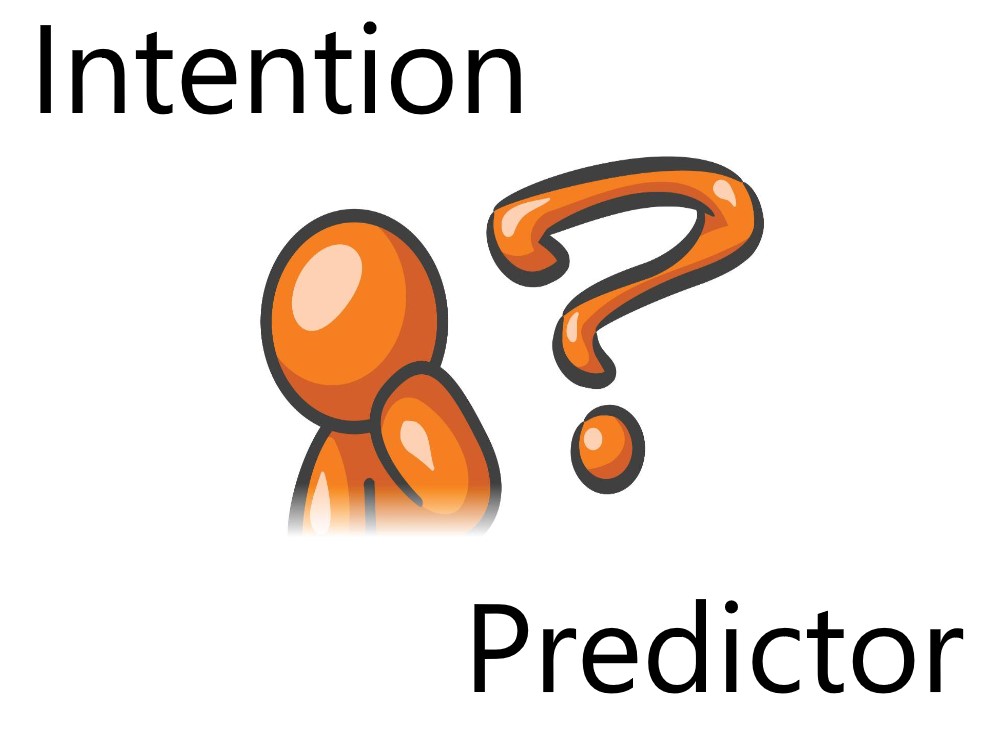
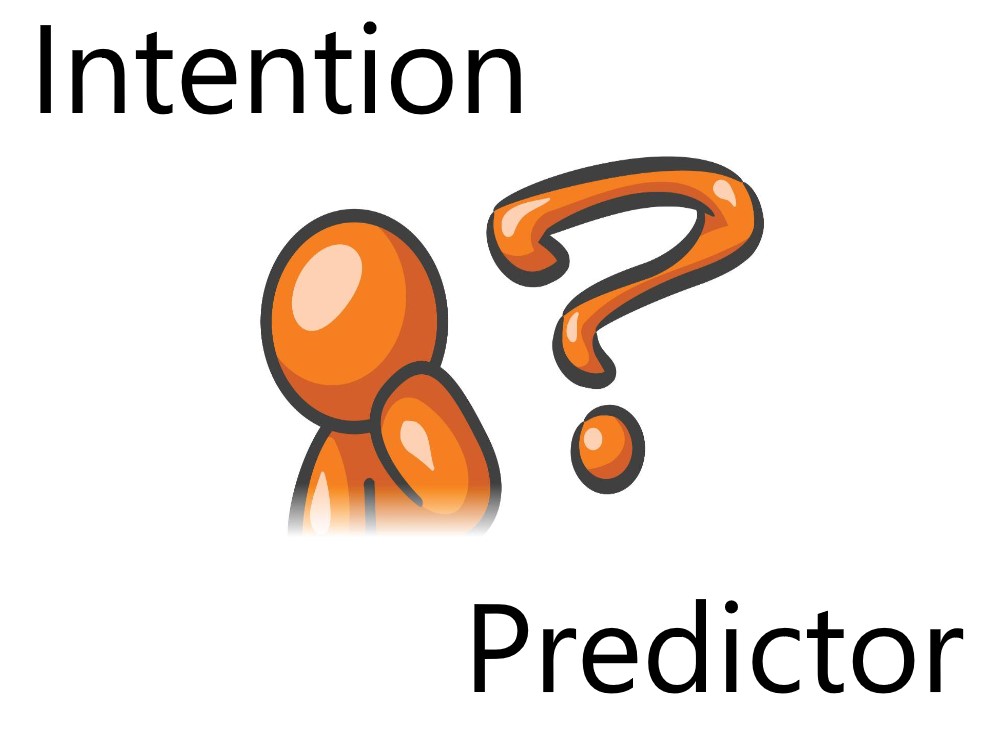
**Consumer Intention Prediction using Twitter**

****

**FINAL YEAR PROJECT (Proposal)**

**Consumer Intention Prediction using Twitter**

****

**Supervisor**

**Miss Nida Saddaf Khan (nskhan@iba.edu.pk)**

**Hasan Imran (himran@khi.iba.edu.pk)**

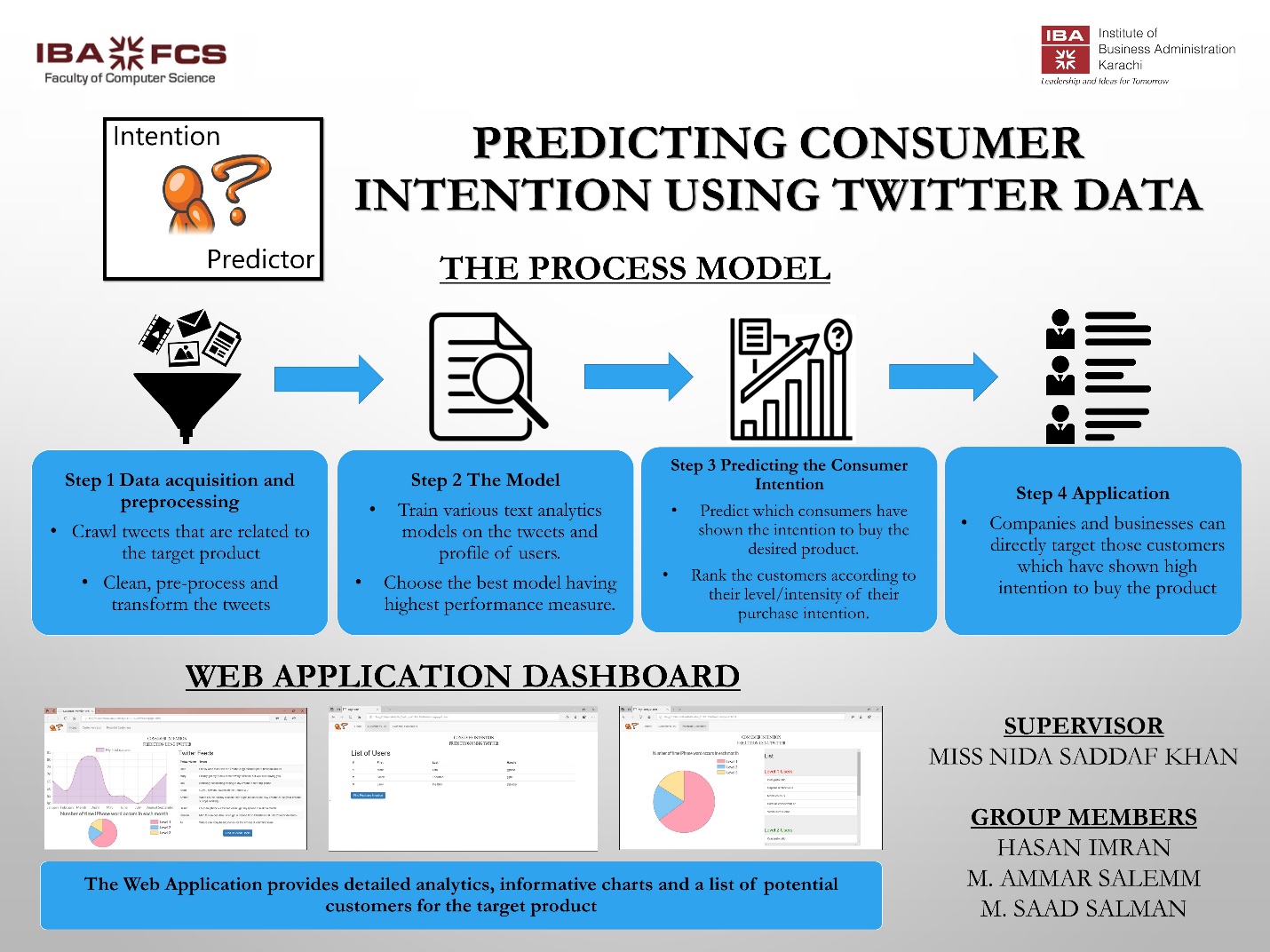
**Muhammad Ammar Saleem (msaleem@khi.iba.edu.pk)**

**Muhammad Saad Salma (muhammadsalman@khi.iba.edu.pk)**

**Executive Summary:**

Our project is a web application that can predict the likelihood/certainty that a customer will buy a product that he is interested in based on his social media posts such as Twitter tweets and user profile data. This will help the company/business target that particular customer more efficiently and boost its sales.

We will be searching for Twitter tweets of potential customers wanting to buy a product. And based on those tweets we will estimate/predict the likelihood that the customer will buy the product. We will make a model by gathering tweets from users who have already expressed intention to buy the product and see their tweet history and if possible, their web search history as well. Using this model, we will input potential customers who have tweeted about the product but have not bought it. And based on the training data the model will estimate a prediction/likelihood of whether the customer will buy it or not. We will limit the scope of our data to only mobile phones. And our model will predict the consumer intention for the latest upcoming mobile phones by gathering data and updating the model at fixed intervals. We plan to generate a trending list of phones on twitter which will be launching soon or have been recently launched and the users can use that list of products on our application to find a customer for that product/mobile phone.





**Project Description:**

1. Background and Motivation

Currently we have many recommendation systems available which recommend different products to the user, most of which are not efficient. No such effective model for businesses to identify potential customers.

We want to develop a software that will help the businesses identify potential customers for their products by estimating their purchase intention in measurable terms from their tweets and user profile data on twitter.

1. Project Goal

We aim to analyze the tweets related to a product and identify the purchase intention in it. In this way we can rank the tweets which have high purchase intention and report the name of the person who tweeted as potential customer of product.

We will make a model by gathering tweets from users who have already expressed intention to buy the product and see their tweet history and if possible, their web search history as well. Using this model, we will input potential customers who have tweeted about the product but have not bought it. And based on the training data the model will estimate a prediction/likelihood of whether the customer will buy it or not.

1. Project Requirements

We require data from twitter to analyze purchase intention. For now, we are gathering the data by scraping the tweets of a product through a scraper. We will need a mechanism to save the scrapped tweets in storage for further processing and performing analysis and for that we have decided to use mongo db. We will need to annotate data retrieved from scrapper. Once model is trained, we will need to develop a website so that users can easily access our application through the graphical interface of the website. We will show on the website the purchase intention rank of tweets on level of 1 to 5 for the desired product of the user.

* 1. Functional Requirements
     1. Website

Website will be required so that users can easily access our application.

* + 1. Dashboard

The user will open the dashboard and see the status of the product through charts and the relevant tweets for the product in a list.

* + 1. Get Relevant Users

The user will press the Get Relevant Users button through which he will be taken to the next screen which will show the list of users and tweets who have shown intention to buy the product.

* + 1. Find Purchase Intention of The Users

The user will then press The Find Purchase Intention Of The Users button through which he will be taken to the final screen which will show the results in a pie chart and list format for all the users who should be targeted and will also show the result in level form meaning that the users will be categorized according the level of intention they have expressed to buy the product.

* 1. Constraints

We are scrapping the data not getting data through twitter Search API. This is sometimes not reliable and robust, so we will need to either get the data from some company which sells such data or will need access to the twitter search API.

Another constraint is that we have a lot of tweets which we will use to train our model and we need annotated data so that we can evaluate our model but since annotating each and every tweet is time-consuming as well as there is no way to verify the annotated tweets, we will have to use methods of majority voting and averaging.

* 1. Objectives

We aim to analyze the tweets related to a product and identify the purchase intention in it.

We want to rank the tweets according to the level of purchase intention and show a list of potential customers to the user which he can use to directly target the customer.

1. Validation and Acceptance Tests

We will test the model accuracy by confusion matrix (Accuracy, Precision, Recall, F-Measure). This will give us a percentage of accuracy achieved by our model.

For our application development, we are going to run unit testing in which individual units of source code, would be tested to determine if they are fit to use. Then, we are going to run integration testing where the individual source codes would be merged and tested as group.

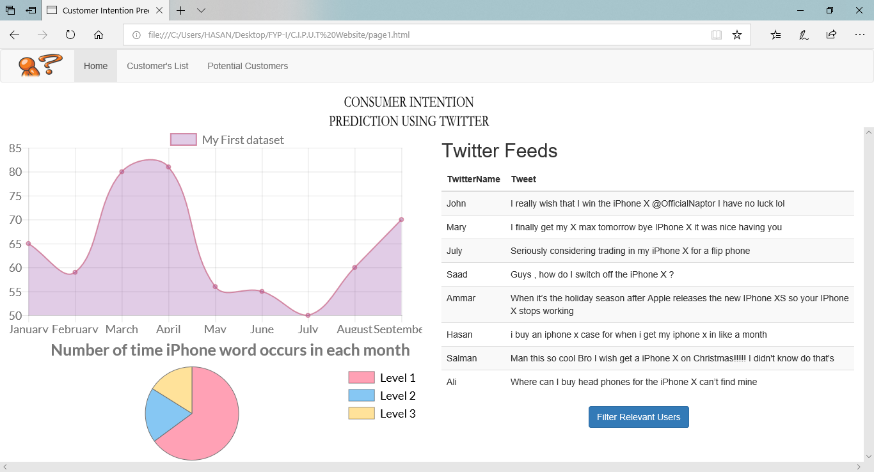
We will also test the usability of our website by carrying out the tasks and functions of the website in different scenarios and checking if they successfully complete.

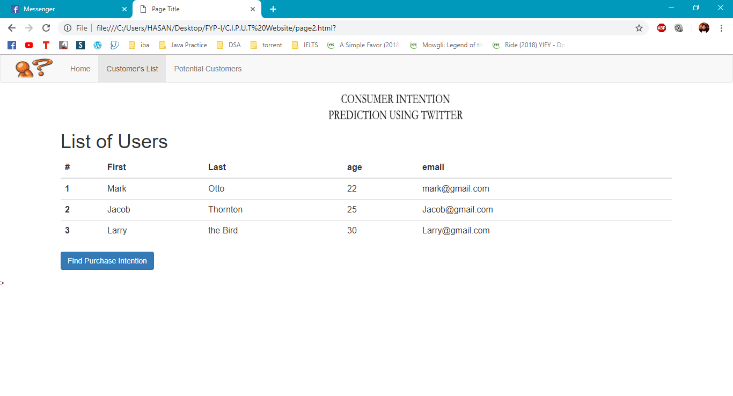
**Technical Design:**

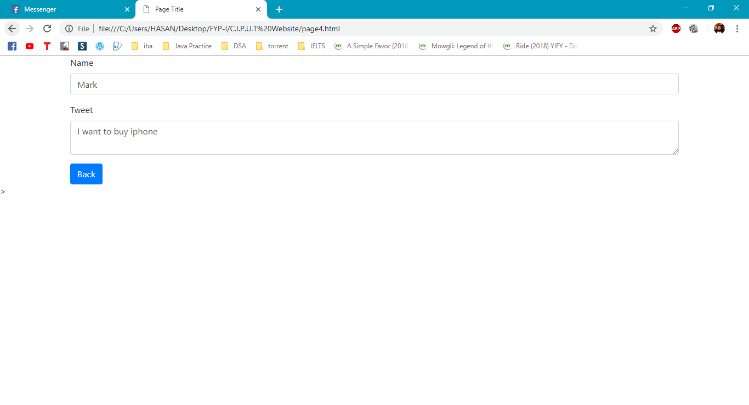
1. Possible Solutions and Design Alternatives

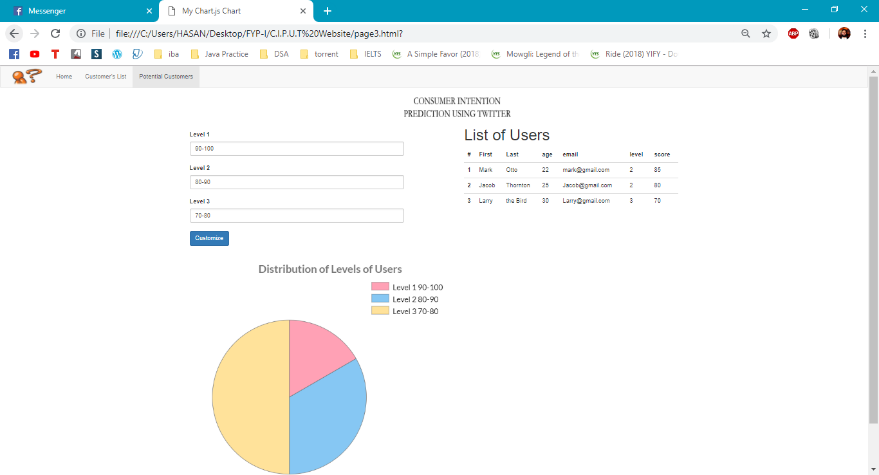
One possible approach is to label the tweets text as having Purchase Intention and Not having Purchase intention. We will collect 30,000 tweets from Twitter using our own web crawler. After removing duplicate tweets, they reserve 15,000 tweets for training data and remaining for testing. We defined definition of Purchase Intention as object that is having action word like (buy, eat) associated with it. We will annotate the data by hiring workers that read their tweets and label them as purchase intention and non-purchase intention tweet. Each tweet was read by 3 people and final class was decided by maximum voting. We will traverse each tweet and find verb which we will categorize as "Purchase Action" words list and list of "Non-Purchase Action" by “pointwise mutual information score ". For feature extraction of post, we will find verb and then find to which category this verb belongs using "WordNet". We will then find "Purchase Object" and "non-Purchase Object" using “pointwise mutual information score ". To extract the Purchase Object words in post we will collect noun phrase and then use "Free base" to categorize each noun phrase as Purchase Object and non-Purchase Object. Then we will find relationship between features with dependency parser, if Consumable Object have Purchase Action word then it is given value as 1 otherwise 0. Then we will find "Purchase Supportive word dependency" to find knowledge about Purchase Object and its corresponding Action word which is found using dependency parser. We will use SVM classifier for training and test the data with annotated tweets and then find accuracy using ROC curve to find accuracy of work.

1. System Level Overview
2. Firstly, the user will open the DASHBOARD and see the status of the product through charts and the relevant tweets for the product in a list.
3. Secondly, the user will press the GET RELEVANT USERS button through which he will be taken to the next screen which will show the list of users and tweets who have shown intention to buy the product.
4. Thirdly, the user will then press the FIND PURCHASE INTENTION OF THE USERS button through which he will be taken to the final screen which will show the results in a pie chart and list format for all the users who should be targeted and will also show the result in level form meaning that the users will be categorized according the level of intention they have expressed to buy the product.









1. Module Level Descriptions
   1. Graphical User Interface Modules

The 1st screen is the home page. It will give a view of the dashboard which contains the status of the product through charts and the relevant tweets for the product in a list.

The 2nd screen is the screen shows the tweets of the relevant customers which have shown purchase intention towards the product.

The 3rd screen contains the final list of customers according to their level of purchase intention shown towards the product so that the user can target them for the desired product.

* 1. Control Modules

Scaping module: the scraping module will display the tweets scraped on the dashboard which contain the product name.

Data processing module: this module will clean and process the tweets which have purchase intention words associated with the desired product.

The machine learning module: this module will use the model built to check if the new tweet has purchase intention, and if so, will calculate the estimated amount of purchase intention for the user of the tweet for the product.

* 1. Miscellaneous Modules

Other modules will be used to generate lists and graphs which we will use to give some amount of analysis easy interpretation of the data for the users.

1. Assessment of Proposed Solution

We will establish different scenarios for our web application. We will check each scenario and the outcome and see if it matches with the expected result.

We will set an accuracy standard for our model and check if the results match the desired standard.

We will also assess the non-functional requirements of the website by evaluating the security, scalability and flexibility of the website in terms of users who use it and for the admins who will monitor the website.

**Work Plan:**

1. Feasibility Assessment

This tool can be used by any organization’s marketing department. This application will help increase the sales of the company as it will target those users who have shown interest in the product. Therefore, it can get positive response in the market.

One of the major concerns of the tool is to input data from different sources, which currently is being done through a scraper which is not reliable in the long term. We will need access to the search API of twitter if we want to reliably gather tweets from the website or we will have to buy the data from companies which sell similar products.

Financial requirements for this tool are minimal because right now we are using a scraper to gather data.

* 1. Skill and Resources

This project requires us to have knowledge about data scraping, data cleaning and processing, applying machine learning models and web application development and design.

* 1. Risk Assessment
* **Technical Risk:** High.
* **Timing Risk:** High.
* **Budget Risk:** Low.

The skills and time are two considerable risks to the project. This is because our goal is to develop a dashboard level website. Any unforeseen technical problem would have a high impact on the progress of the project.