**Artificial Intelligence**

**Project Proposal**

**Project title:**

Sentiment Analysis on Android/IOS Applications

**Submitted by:**

Abdul Hameed Khan (FA18-BCS-002)

Daud Ahmed (FA18-BCS-010)

**Submitted to:**

Ma’am Ayesha Jadoon

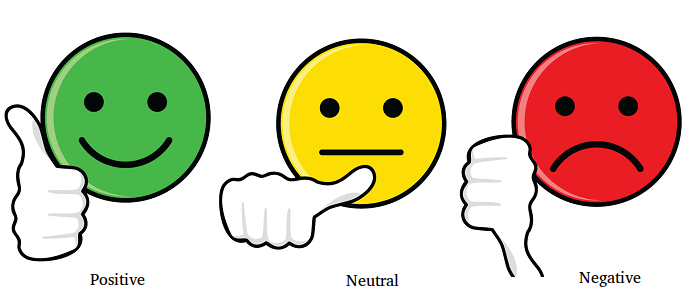
**Introduction:**

As we know that if any developer wants to know about his application on play store or app store then he needs to check all the comments/reviews that he received on his application. There is no any application for this problem on Play store and App store for giving us sentiment about our application; we have to manually check all the reviews. Therefore, it is very difficult for human to do it manually because sometime application having thousands or millions of Reviews so it is impossible for human to do sentiment of that app and there is millions of applications on Play Store and App store. The users of the software and apps present in the android stores and other online providers give their reviews about the products so it is very difficult for the developers to go through all the reviews given by the customer or user.

**What is Sentiment Analysis?**

“Sentiment Analysis is the computational evaluation of documents to determine the fine-grained emotions that are expressed.”[1]

Sentiment analysis is a kind of data mining where you measure the inclination of people’s opinions by using NLP (natural language processing), text analysis, and computational linguistics. We perform sentiment analysis mostly on public reviews, social media platforms, and similar sites. [2]



We all use different apps of play stores and app store and we give different opinions on different apps. There is huge amount of reviews on every app. Now if the developer of an app wants to know about his app that whether the people has given positive reviews, negative reviews and neutral on his app.

Now the question is how a developer will analyze his app when the reviews [data] is in huge amount like thousands of reviews on every app. Here a problem takes place that how a developer will know about his app.

Here comes sentiment analysis when no traditional way can solve this problem. The sentiment analysis uses machine learning and Natural language processing (NLP) to analyze such a vast majority of reviews or comments and provide better results for this task.

**Background information**

The origin of sentiment analysis can be traced to the 1950s, when sentiment analysis was primarily used on written paper documents. Today, sentiment analysis is widely used to extract subjective information from content on the Internet, including texts, tweets, blogs, social media, news articles, reviews, and comments. This is done using a variety of different techniques, including NLP, statistics, and machine learning methods. Organizations then use the information mined to identify new opportunities and better target their message toward their target demographics.

**Benefits:**

1. Upselling opportunities
2. Agent monitoring
3. Training chatbots
4. Identifying key emotional triggers
5. Handling multiple customers
6. Adaptive customer service
7. identify a dissatisfied customer
8. [customer satisfaction](https://www.whoson.com/customer-service/3-simple-steps-to-boost-live-chat-customer-satisfaction/)

**Methodology:**

**SVM (Support Vector Machine)**

“Support Vector Machine” (SVM) is a supervised machine learning algorithm which can be used for either classification or regression challenges. However, it is mostly used in classification problems. In the SVM algorithm, we plot each data item as a point in n-dimensional space (where n is number of features you have) with the value of each feature being the value of a particular coordinate. Then, we perform classification by finding the hyper-plane that differentiates the two classes very well (look at the below snapshot).

[](https://www.analyticsvidhya.com/wp-content/uploads/2015/10/SVM_1.png)

Support Vectors are simply the co-ordinates of individual observation. The SVM classifier is a frontier, which best segregates the two classes (hyper-plane/ line).

You can look at support vector machines and a few examples of its working here.

**How does it work?**

* **Identify the right hyper-plane (Scenario-1):**Here, we have three hyper-planes (A, B and C). Now, identify the right hyper-plane to classify star and circle.  
  You need to remember a thumb rule to identify the right hyper-plane: “Select the hyper-plane which segregates the two classes better”. In this scenario, hyper-plane “B” has excellently performed this job.0
* **Identify the right hyper-plane (Scenario-2):**Here, we have three hyper-planes (A, B and C) and all are segregating the classes well. Now, How can we identify the right hyper-plane?

Here, maximizing the distances between nearest data point (either class) and hyper-plane will help us to decide the right hyper-plane. This distance is called as **Margin**. Let’s look at the below snapshot:[[](https://www.analyticsvidhya.com/wp-content/uploads/2015/10/SVM_4.png)](https://www.analyticsvidhya.com/wp-content/uploads/2015/10/SVM_4.png)Above, you can see that the margin for hyper-plane C is high as compared to both A and B. Hence, we name the right hyper-plane as C. Another lightning reason for selecting the hyper-plane with higher margin is robustness. If we select a hyper-plane having low margin then there is high chance of miss-classification.

* **Identify the right hyper-plane (Scenario-3):**Hint:Use the rules as discussed in previous section to identify the right hyper-plane

**[](https://www.analyticsvidhya.com/wp-content/uploads/2015/10/SVM_5.png)**Some of you may have selected the hyper-plane **B**as it has higher margin compared to **A.**But, here is the catch, SVM selects the hyper-plane which classifies the classes accurately prior to maximizing margin. Here, hyper-plane B has a classification error and A has classified all correctly. Therefore, the right hyper-plane is **A.**

* **Can we classify two classes (Scenario-4)?:**Below, I am unable to segregate the two classes using a straight line, as one of the stars lies in the territory of other(circle) class as an outlier.  **[](https://www.analyticsvidhya.com/wp-content/uploads/2015/10/SVM_61.png)**As I have already mentioned, one star at other end is like an outlier for star class. The SVM algorithm has a feature to ignore outliers and find the hyper-plane that has the maximum margin. Hence, we can say, SVM classification is robust to outliers.  
  **[](https://www.analyticsvidhya.com/wp-content/uploads/2015/10/SVM_71.png)**
* **Find the hyper-plane to segregate to classes (Scenario-5):**In the scenario below, we can’t have linear hyper-plane between the two classes, so how does SVM classify these two classes? Till now, we have only looked at the linear hyper-plane.**[](https://www.analyticsvidhya.com/wp-content/uploads/2015/10/SVM_8.png)**

SVM can solve this problem. Easily! It solves this problem by introducing additional feature. Here, we will add a new feature z=x^2+y^2. Now, let’s plot the data points on axis x and z:  
[[](https://www.analyticsvidhya.com/wp-content/uploads/2015/10/SVM_9.png)](https://www.analyticsvidhya.com/wp-content/uploads/2015/10/SVM_9.png)In above plot, points to consider are:

* + All values for z would be positive always because z is the squared sum of both x and y
  + In the original plot, red circles appear close to the origin of x and y axes, leading to lower value of z and star relatively away from the origin result to higher value of z.

In the SVM classifier, it is easy to have a linear hyper-plane between these two classes. But, another burning question which arises is, should we need to add this feature manually to have a hyper-plane. No, the SVM  algorithm has a technique called the **kernel trick**. The SVM kernel is a function that takes low dimensional input space and transforms it to a higher dimensional space i.e. it converts not separable problem to separable problem. It is mostly useful in non-linear separation problem. Simply put, it does some extremely complex data transformations, then finds out the process to separate the data based on the labels or outputs you’ve defined.

When we look at the hyper-plane in original input space it looks like a circle:  
[](https://www.analyticsvidhya.com/wp-content/uploads/2015/10/SVM_10.png)

Now, let’s look at the methods to apply SVM classifier algorithm in a data science challenge.

**Functional Requirements:**

1. The system can take input the app Id form the user, then the system should be able to search the app id on the play store or app store.
2. After searching, the system should be able to get all the reviews of the app from play store or app store and can make a csv file in which this data will be stored.
3. The system can analyze the dataset and can perform the classification in terms of positive, negative, neutral etc.
4. The system can perform the classifier training process and display the model in the form of feature sets of the term data from the training data.
5. The system can display the test data result and display the results in the form of graphs generated from the classifier testing.
6. The system can display sentiment analysis result derived from reviews submitted by users.

**2.9 Project Approach**

In this system we will use Machine learning and Natural Language Processing (NLP) to classify sentiments on reviews of app.

Firstly, we need a dataset to make our ML model and then train the data set on this model. We can get this dataset by scraping technique, which I will not explaining here. When our ML model will train then we will test its accuracy. If the accuracy is satisfying then we will use it further but if the accuracy is not satisfying then we will train it again. This process will repeat until we get the high accuracy.

Now on the front end the user will give input its app id and check for results. At the backend, we will scrap its app’s reviews from the play store or app store through google play store API. These scraped reviews will be test on our model and then our model will show results according to the given data.

The results will show as graphs, tables, every review, and its sentiment will be label with it.

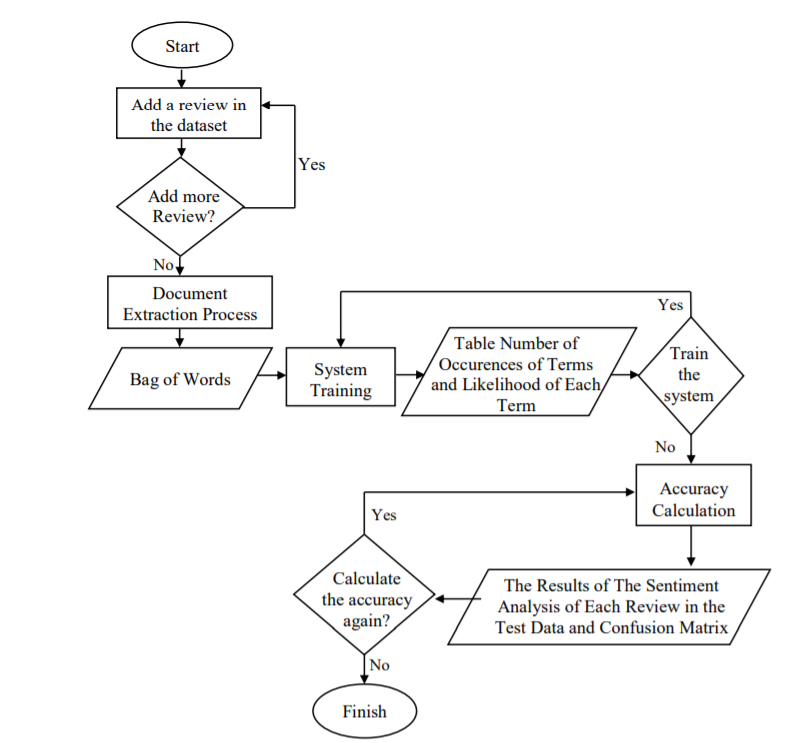


Fig: Flowchart of App review sentiment analysis system

**Stakeholders:**

* Customers / Clients

Our target audience will be the developers, companies, organizations who will use our product according to their requirements.

* Employees

The team who will make this software, which include designers, coders, testers etc.

* Investors

Definitely, in every project we need some investments and support from the people who can help to make our software successful. Investors include both shareholders and debtholders. Shareholders invest capital in the business and expect to earn a certain rate of return on that invested capital. Investors are commonly concerned with the concept of shareholder value. [3]

**2.5 Scope, Impact and interdependencies:**

**Scope and Impact:**

Sentiment analysis provides an environment and platform for the developer to easily go through the reviews given by the user also sentimental analysis is being used by many organizations to get reviews about their products.

**Interdependencies:**

Sentiment analysis depends on the reviews, reviews depends upon the users of the product, users depend upon the product, product depend upon the organization or developer and organization is independent.

**References:**

[1] “Sentiment Analysis: Beyond Polarity”, https://www.cs.bham.ac.uk/~smithpm/publications/documents/rsmg3.pdf

[2] “Sentiment Analysis Projects & Topics For Beginners [2021]”, https://www.upgrad.com/blog/sentiment-analysis-projects-topics-for-beginners/