A PROJECT SYNOPSIS

On

“Dynamic Product Review”

**Submitted for Partial Fulfilment of Award of**

**BACHELOR OF TECHNOLOGY**

**In**

**Computer Science & Engineering**

**Subject Name & Code: Project (KCS 753)**

Submitted By:

Aditya Kumar Yadav (1905080100006)

Aditya Jayant Gupta (1905080100005)

Harsh Prajapati (1905080100035)

### Under the Guidance of:

Dr. Avinash Gupta (Head of Department)



**BABU BANARASI DAS ENGINEERING COLLEGE, LUCKNOW**

**Affiliated to**

**Dr. APJ ABDUL KALAM TECHNICAL UNIVERSITY, LUCKNOW**

(Semester: - VIIth , 2022-23)

**DECLARATION**

I hereby declare that the Major Project entitled **("Dynamic product review")**is an authentic record of my own work as Major Project for the award of degree of B.Tech. (Computer Science and Engineering). Babu Banarsi Das Engineering College, Lucknow, under the guidance of (Dr. Avinash Gupta).

Aditya Jayant Gupta (1905080100005)

Date:- Harsh Prajapati (1905080100035)

Aditya Kumar Yadav (1905080100006)

# TABLEOFCONTENT

1. [Introduction 1-2](#_TOC_250021)
   1. Introduction to Disease Prediction 1
   2. Literature Survey 2
2. [Problem Definition 3-4](#_TOC_250020)
   1. [Project Objective 3](#_TOC_250019)
   2. [Proposed Methodology 4](#_TOC_250018)
3. [Technology Used 5](#_TOC_250017)
   1. [Sentiment Analysis 5](#_TOC_250016)
   2. [Web Scrapper 5](#_TOC_250015)
4. Algorithm Used 6-9
   1. [Logistic Regression 6](#_TOC_250014)
   2. [Naïve Bayes Classifier 7](#_TOC_250013)
   3. [Support Vector Machine 8](#_TOC_250012)

[4.4 Random Forest.......................................................................](#_TOC_250011)

1. [Mathematical Representation 10-11](#_TOC_250010)
   1. [Precision 10](#_TOC_250009)
   2. [Recall 10-11](#_TOC_250008)
   3. [F1 Score 11](#_TOC_250007)
2. [Technological Requirements 12](#_TOC_250006)
   1. [Hardware Requirements 12](#_TOC_250005)
   2. [Software Requirements 12](#_TOC_250004)
3. [Module Description 13-14](#_TOC_250003)
4. [Data Flow Diagram 15-16](#_TOC_250002)
5. Advantages/Application/Limitations 17
   1. Applications /Advantages 17
   2. [Limitations 17](#_TOC_250001)
6. [References 18](#_TOC_250000)

# INTRODUCTION

This project is titled “Dynamic Product Review Using Sentiment Analysis”. It is general idea of searching any given product by user on the basis of comments/reviews by user who have already used the product or have ordered from that specific site. Only on the basis of ratings it is not possible to judge the product that it is good or bad.

In this project we are going to search the product from two sites and then one the basis of ratio of good or bad comments we will show the result to the user from both the sites. We will use Web scrapping to get the detail of the desired products from both the sites and the convert it in a form so that we can separate out the main required data, and after that with the help of ML with python we will judge the sentiment of comments and then we will show the result to user on the basis of ratio of good and bad comments. We have taken name Dynamic product review because every time you search for the any specific product we are going to search again and do the whole process again and after that it is possible that the order of the product get changed, So basically the result will be dynamic.

The aim of this project is to identify and analyze the probable sentiments of reviews for given products across shopping platforms and provide a visual representation of the customers’ feelings towards said product.

This project was chosen because we have seen that many user buy the product and are not satisfied by the product many times and because of these basic problems:-

* It is seen that in the context of online shopping, most people typically buy products based off of the general product ratings.
* While ratings provide a general idea, they are not the best indicator of the customer’s experience with the product.
* Most people find reading through reviews cumbersome and exhausting. Therefore, by giving a brief, general idea of the sentiments left by the customers in product reviews, we can enable people to make better decisions.

## LITERATURE SURVEY:-

There have been numerous studies on dynamic product reviews using sentiment analysis and text summarization. One early study by Pang and Lee (2008) applied machine learning techniques to movie reviews and found that the use of sentiment analysis can improve the accuracy of the classification of positive and negative reviews. Another study by Hu and Liu (2004) proposed a framework for opinion mining that combines sentiment analysis with text summarization to extract subjective information from online reviews.

More recent studies have focused on the use of deep learning techniques for sentiment analysis and text summarization. For example, a study by Koyuncu and Alpaydin (2018) applied deep learning to Turkish hotel reviews and found that the use of sentiment analysis and text summarization can improve the accuracy of the classification of positive and negative reviews.

Other studies have focused on the integration of sentiment analysis and text summarization in practical applications, such as product recommendation systems. A study by Zhang et al. (2017) proposed a recommendation system that uses sentiment analysis and text summarization to generate personalized product recommendations based on the sentiments expressed in the reviews.

The current methods for determining the product quality and consumer experience rely heavily on lists of reviews and simple average star numbers. While simple and easy to comprehend, these methods present their own set of problems including but not limited to - reviews mismatching their star ratings, too much text to go through, limited searchability and so on.

Major brands have started employing sentiment analysis on their product reviews but most only use it for internal algorithmic tuning and not customer facing interfaces. The existing techniques for product review analysis are listed below in table 1 and table 2.

Static methods of analysis lack scalability and modularity. While they may suit certain niche use-cases, for more dynamic ones they cannot be employed without extensive manual input. Dynamic methods of analysis exist in the form of various models and enterprise solutions. The commercial solutions are expensive and intended for organizational use for the most part. They are too costly and too complex for the general user. Models using SVMs are good to train on simpler classification tasks but fail when it comes to handling long distance contexts as encountered in review analysis. Logistical regression models can be applied but only when dealing with binary classification of positive and negative and that too with limited to no context awareness. Among the various models, transformer based models have proved to be the most accurate and easiest to work with.

Overall, the literature on dynamic product reviews using sentiment analysis and text summarization suggests that these techniques can be effective in extracting useful information from large volumes of text and can be applied in practical applications such as recommendation systems. However, there are still challenges to be addressed, such as the need for large amounts of labeled data and the difficulty of handling subjective and context-dependent language.

# PROBLEM DEFINITION

In this modern era, day to day life became smarter and interlinked with technology. We already know some of the sentiment analysis platform such as bot at twitter the main aim of this sentiment analysis is to judge the mood of user that what will be emotion of the user who has written that comment i.e., is he happy, sad, unsatisfied, etc. So we are going to use that sentiment analysis and we will judge what amount of user are satisfied by some specific product and what amount of people are unhappy from that product that will give us a general idea about the product , is that product quality good on that seller on that site or some other seller is giving better quality of the same product . It is not possible to judge the quality of the product only on the basis of ratings so we will use sentiment analysis to find the good vs. bad comment on a product and then judge that product

## PROJECT OBJECTIVE

* **Identify and analyze the possible sentiments of reviews for the given products**:- To find out what is the emotion of the given the product i.e the comment which we are reading is a good comment or bad the user who have written that comment is he happy with the product or not.
* **Create a web application frontend to allow users to search for a product:-** We will create a web application with the help of python which will be visible to the user and he/she can write down the product they want to find out.
* **Provide a visual representation of the generated analysis data:-**
* **Enable searching across the Amazon and Flipkart shopping platforms**:- one of the main objective is to find the desired product from more than just one site i.e we will get the result from two different sites and then we will carry out the process of judging the product and then from the results of both the sites we will decide which product should be recommended first.

## PROPOSED METHODOLOGY

* Create a web frontend to allow users to search for products
* Set up a custom sentiment analysis API that combines the results from various pre-trained available SA models
* Create a web scraper to fetch product review information from Amazon and Flipkart websites
* Once processed, store the generated sentiment report in a frequently updating cache
* Visualize the review data on frontend using JavaScript visualization libraries.

# TECHNOLOGY USED

## Sentiment Analysis

Sentiment analysis, also referred to as opinion mining, is an approach to natural language processing ([NLP](https://www.techtarget.com/searchbusinessanalytics/definition/natural-language-processing-NLP)) that identifies the emotional tone behind a body of text. This is a popular way for organizations to determine and categorize opinions about a product, service, or idea. It involves the use of [data mining](https://searchsqlserver.techtarget.com/definition/data-mining), machine learning ([ML](https://www.techtarget.com/searchenterpriseai/definition/machine-learning-ML)) and artificial intelligence ([AI](https://www.techtarget.com/searchenterpriseai/definition/AI-Artificial-Intelligence)) to [mine text](https://www.techtarget.com/searchbusinessanalytics/definition/text-mining) for sentiment and subjective information.

Sentiment analysis systems help organizations gather insights from unorganized and [unstructured text](https://www.techtarget.com/searchbusinessanalytics/definition/unstructured-text) that comes from online sources such as emails, blog posts, support tickets, web chats, social media channels, forums and comments. Algorithms replace manual data processing by implementing rule-based, automatic or hybrid methods. Rule-based systems perform sentiment analysis based on predefined, lexicon-based rules while automatic systems learn from data with machine learning techniques. A hybrid sentiment analysis combines both approaches.

In addition to identifying sentiment, opinion mining can extract the polarity (or the amount of positivity and negativity), subject and opinion holder within the text. Furthermore, sentiment analysis can be applied to varying scopes such as document, paragraph, sentence and sub-sentence levels.

## Text Summarization

Text summarization is the process of generating a concise and coherent summary of a larger text, and it is often used to reduce the time and effort required to read and understand large volumes of text. There are several approaches to text summarization, including rule-based methods and machine learning-based methods. Machine learning-based text summarization techniques can be classified into two main categories: extractive summarization and abstractive summarization..

Overall, machine learning-based text summarization techniques have the potential to significantly reduce the time and effort required to read and understand large volumes of text, but there are still challenges to be addressed, such as the need for large amounts of labeled data and the difficulty of handling subjective and context-dependent language.

## Web Scrapper

Web scraping is an automatic method to obtain large amounts of data from websites. Most of this data is unstructured data in an HTML format which is then converted into structured data in a spreadsheet or a database so that it can be used in various applications. There are many different ways to perform web scraping to obtain data from websites. These include using online services, particular API’s or even creating your code for web scraping from scratch. Many large websites, like Google, Twitter, Facebook, StackOverflow, etc. have API’s that allow you to access their data in a structured format. This is the best option, but there are other sites that don’t allow users to access large amounts of data in a structured form or they are simply not that technologically advanced. In that situation, it’s best to use Web Scraping to scrape the website for data.Web scraping requires two parts, namely the **crawler** and the **scraper**. The crawler is an artificial intelligence algorithm that browses the web to search for the particular data required by following the links across the internet. The scraper, on the other hand, is a specific tool created to extract data from the website. The design of the scraper can vary greatly according to the complexity and scope of the project so that it can quickly and accurately extract the data.

# ALGORITHMS USED

## Transformer Neural Networks

Transformer neural networks are a type of deep learning architecture that has been widely used in natural language processing (NLP) tasks, such as machine translation, language modeling, and text summarization. They were first introduced by Vaswani et al. (2017) and have since become one of the most popular and effective models for NLP tasks.

One of the key features of transformer neural networks is the use of self-attention mechanisms, which allow the model to consider the relationships between all input elements (e.g. words in a sentence) simultaneously rather than processing them sequentially. This allows the model to capture long-range dependencies in the input data, which is important for tasks such as machine translation, where the meaning of a word can depend on the context of words that are far away in the sentence.

Another key feature of transformer neural networks is the use of multi-headed attention, which allows the model to attend to different parts of the input data simultaneously. This allows the model to learn multiple different relationships between the input elements, which can improve its ability to understand the meaning and context of the input data.

In addition to self-attention mechanisms and multi-headed attention, transformer neural networks also use feed-forward layers and residual connections, which allow the model to learn more complex patterns in the data and improve its ability to generalize to new data.

Overall, transformer neural networks have proven to be highly effective for a wide range of NLP tasks, and they have significantly improved the state-of-the-art in many areas. However, they do have some limitations, such as the need for large amounts of labeled data and the difficulty of training on large input sequences.

## BERT Transformer model

BERT (Bidirectional Encoder Representations from Transformers) is a transformer neural network model developed by Google for natural language processing (NLP) tasks. It was introduced by Devlin et al. (2018) and has become one of the most widely used and successful models in NLP, achieving state-of-the-art results on a wide range of tasks.

One of the key features of BERT is that it is a bidirectional model, meaning that it processes the input data in both the forward and backward directions. This allows the model to consider the context of both the preceding and following words when making predictions, which is important for tasks such as language modeling and machine translation, where the meaning of a word can depend on the context of words that are far away in the sentence.

Another key feature of BERT is its use of transformer architecture, which allows the model to use self-attention mechanisms to consider the relationships between all input elements (e.g. words in a sentence) simultaneously rather than processing them sequentially. This allows the model to capture long-range dependencies in the input data, which is important for understanding the context and meaning of the input data.

BERT has been trained on a large dataset of unlabeled text and can be fine-tuned for specific tasks by adding a task-specific layer on top of the pre-trained model. This allows BERT to be used for a wide range of NLP tasks, including language modeling, machine translation, text classification, and question answering.

Overall, BERT has proven to be a highly effective model for NLP tasks and has significantly improved the state-of-the-art in many areas. However, it does have some limitations, such as the need for large amounts of labeled data and the difficulty of training on large input sequences.

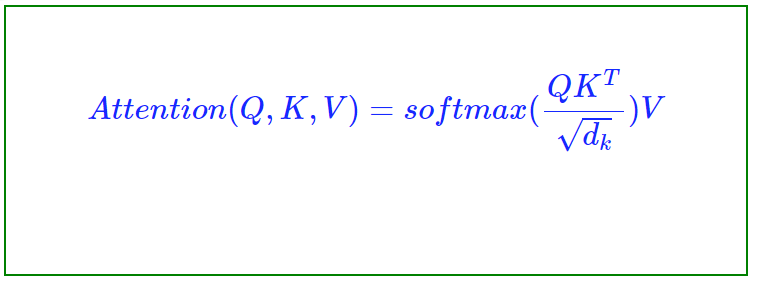
# MATHEMATICAL REPRESENTATION

## Transformers

The mathematical representation of transformers, a type of deep learning architecture used in natural language processing (NLP) tasks, is based on the use of self-attention mechanisms and feed-forward layers.

In a transformer model, the input data is first transformed into a sequence of vectors, where each vector represents a word or a sequence of words (e.g. a subword). These vectors are then fed into the self-attention mechanism, which computes a weighted sum of the vectors based on their relationships with each other.

Mathematically, the self-attention mechanism can be represented as follows:



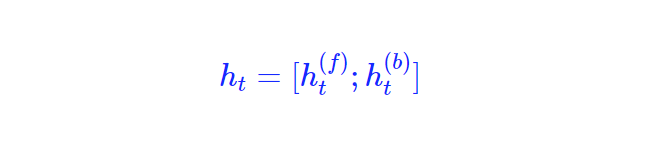
Where Q, K, and V are matrices representing the query, key, and value vectors, respectively, and $d\_k$ is the dimensionality of the key vectors. The self-attention mechanism computes the dot product between the query and key vectors and divides it by the square root of the dimensionality of the key vectors in order to scale the dot product. The result is then passed through a softmax function to compute the attention weights, which are used to weigh the value vectors and compute the weighted sum.

The output of the self-attention mechanism is then passed through a feed-forward layer, which consists of a linear transformation followed by a non-linear activation function (e.g. ReLU). The feed-forward layer is used to learn more complex patterns in the data and improve the model's ability to generalize to new data.

**5.2 Bi-directional processing mechanism**

The bidirectional processing mechanism is a technique used in deep learning models, such as the BERT (Bidirectional Encoder Representations from Transformers) model, to consider the context of both preceding and following words when making predictions.

Mathematically, the bidirectional processing mechanism can be represented as follows:

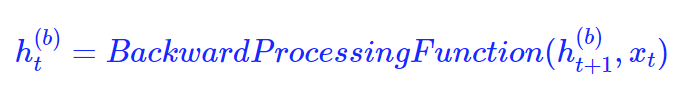


Where ht(f) and ht(b) are the forward and backward hidden states at time step t, respectively, and [;] is the concatenation operator.

To compute the forward hidden states, the input data is processed in the forward direction using a forward processing function, such as a recurrent neural network (RNN) or a transformer:



To compute the backward hidden states, the input data is processed in the backward direction using a backward processing function, such as a backward RNN or a transformer:



# TECHNOLOGICALREQUIREMENTS

## HARDWARE REQUIREMENTS

* + 1. **Processor -** intelCOREi3
    2. **HardDisk - 10** GB
    3. **Monitor -** 15 VGAColor
    4. **RAM -** 4GB
    5. **Mouse -** Optical
    6. **Keyboard -** Multimedia

## SOFTWARE REQUIREMENTS

* Python 3
* Visual Studio Code
* Huggingface API for interacting with transformer models
* Heroku/Google Cloud/Microsoft Azure for hosting
* Firebase and Firestore for authentication and data storage

# MODULE DESCRIPTION

## DATASET

Thismodule deals with thebuilding of the dataset on the basis of which our modelwill be trained. Applying Knowledge to field of Medical Science and making the taskof Physician easy is the main purpose of this dataset. Currently in this project, threedatasetswillbeusedfordiabetes,heart,and parkinson’sdiseaserespectively.

## PRE-PROCESSING

Preparing raw data to be used with a machine learning model is known as data pre- processing.In order tobuildamachinelearningmodel,itis thefirstandmostimportant stage. It is not always the case that we come across the clean and prepareddata when developing a machine learning project. Therefore, the most crucial stage ofa machine learning project is pre-processing (cleaning). The quality of our machinelearningmodelisdeterminedonthequalityofourdata.Therefore,cleaningthedatais alwaysrequiredbeforefeedingittothemodelfortraining.

* 1. **CLASSIFICATION** Itispossibletodoclassificationonbothstructuredandunstructureddata.Classification is the act of dividing a given set of data into classes. Predicting the classof the provided data points is the first step in the procedure. The terms target, label,and classes are frequently used to describe the classes. The task of approximating themappingfunctionfromdiscreteinputvariablestooutputvariablesisknownasclassificati on predictivemodelling. Finding the class or category that the new datawillbelongtois thekeyobjective.

## APPLY KNOWLEDGE

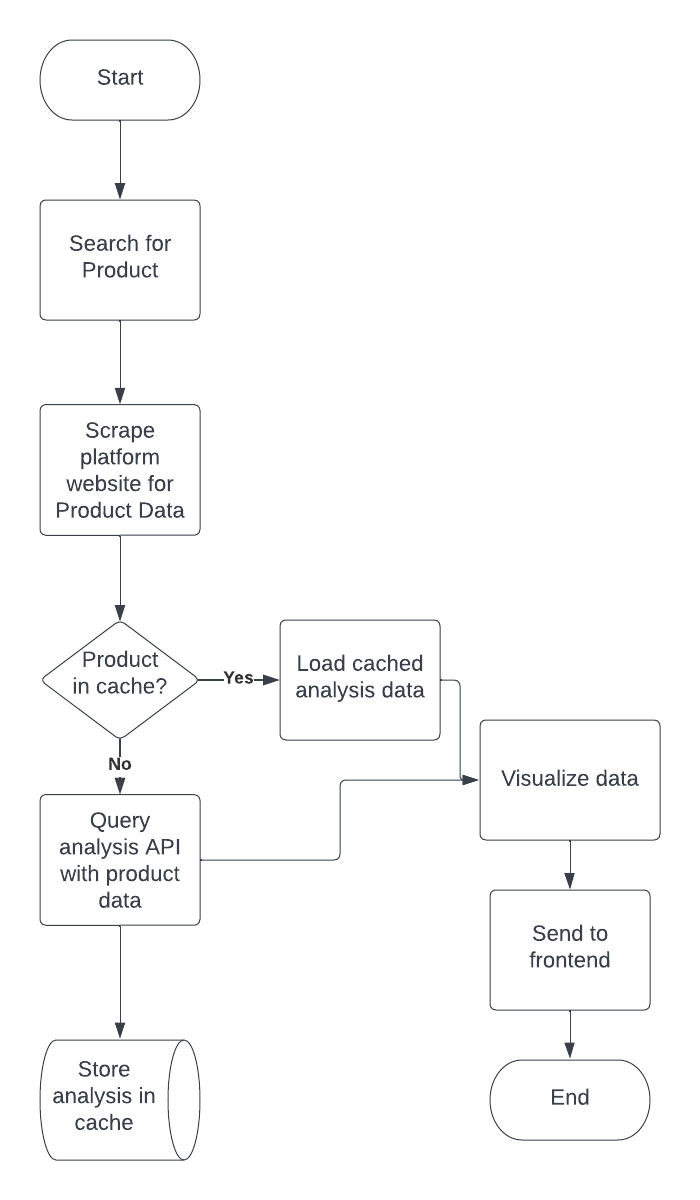
In several domains, machine learning has been extensively studied and used. Highprecision needs lot of data, though, and getting that data can occasionally bechallenging,expensive,orunfeasible.Byincorporatinghumanexpertiseintomachine

learning,onecandrasticallyminimizetheamount ofdataneeded,improvetheaccuracyand resilience of the technology, and create machine learning systems that are easy tounderstand. This will enable interaction between people and machine learning systems,making machine learning conclusions understandable to humans, and will allow usingthe immense quantity of human knowledge and machine learning potential to reachfunctionalitiesandperformancenotpreviouslyattainable.

## PREDICTEDOUTPUTMODEL

We will be able to forecast the disease from the input symptoms using a single modelor by integrating the predictions of many models after training one or more models.This strengthens and improves the accuracy of our total prediction. Finally, we willdefine a function that accepts a listof symptoms separated by commas, uses thetrained models to predict the disease from the symptoms, and outputs the predictionsinalistorvectorformat.

# DATA FLOW DIAGRAM

****

**Fig7.1DataFlowDiagram**

Theprogressoftheuserwiththewebsitestartswithaccessingthelinkofwebpage.Thiswil lallowthemtoaccessthe website.

Afterreachingthewebsite,thefirstpagedisplayedonthescreenwillbetheDiabetesPredictio n, on the left side of the page there is an option menu from where user canselect the disease for which he/she wants to perform prediction by clicking on thename ofthatdisease.

Afterreachingthepageofparticulardiseaseprediction, userneedstofeedsomeattributesorsymptomsrelatedtothatdiseaseforprediction.

Afterfeeding thesymptoms,ourmodelwillperformcomputationandcomputetheresult.

Theresultwilldisplayedonthescreen,indicatingwhethertheuserhavethatdiseaseornot,afte rperformingcomputation.

# ADVANTAGES/LIMITATIONS

## ADVANTAGES/APPLICATIONS

Withanincreaseinbiomedicalandhealthcaredata,accurateanalysisofmedicalda ta benefits earlydiseasedetectionandpatientcare.

Optimizeoperationsand proactivecare.

Determineoverallcohorthealth.

Highqualityand patient-centeredservices.

Bettertreatmenttargetingandriskprediction.

## LIMITATIONS

Timecomplexityismore,increaseswhilesearchingforinsignificantbranchesandlast lynoprecautionsaredefined.

If attributes are not related then Decision trees prediction is less accurate andANNiscomputationallyintensivetotrainalsoitdoesnotleadtospecificconclusion.

# REFERENCES

[1] Fang, X., Zhan, J. Sentiment analysis using product review data. Journal of Big Data 2, 5 (2015). **<https://doi.org/10.1186/s40537-015-0015->**2

[2] NLP and Sentiment analysis by symbl.ai, **<https://symbl.ai/about>**

[3] Tamir Bennatan, Conducting Sentiment Analysis on App Reviews to Inform Product Decisions. Hopper, Medium (Jun 13, 2019). **<https://medium.com/life-at-hopper/conducting-sentiment-analysis-on-app-reviews-to-inform-product-decisions-64fcc71822ed>**

[4] **[Atharva Mashalkar](https://atharva-mashalkar.medium.com/?source=post_page-----16b806eb4c4b--------------------------------)**, Sentiment Analysis using Logistic Regression and Naive Bayes. TowardsDataScience (Nov 28, 2020). **<https://towardsdatascience.com/sentiment-analysis-using-logistic-regression-and-naive-bayes-16b806eb4c4b>**

[5] Thura, T., Na, J.-C., & Khoo, C. S. G. (2010). Aspect-based sentiment analysis of movie reviews on discussion boards. Journal of Information Science, 36(6), 823–848. **<https://doi.org/10.1177/0165551510388123>**

[6] **[Hu Xu](https://arxiv.org/search/cs?searchtype=author&query=Xu,+H)**, **[Bing Liu](https://arxiv.org/search/cs?searchtype=author&query=Liu,+B)**, **[Lei Shu](https://arxiv.org/search/cs?searchtype=author&query=Shu,+L)**, **[Philip S. Yu](https://arxiv.org/search/cs?searchtype=author&query=Yu,+P+S)** (2019). BERT Post-Training for Review Reading Comprehension and Aspect-based Sentiment Analysis. **<https://doi.org/10.48550/arXiv.1904.02232>**

1. [7] Federico Pascual, Analyze Sentiment in Product Reviews. MonkeyLearn Blog (Mar 23, 2019). **<https://monkeylearn.com/blog/sentiment-analysis-of-product-reviews/>**