# **C:\Users\Anes\Pictures\Saved Pictures\burchLOGO.jpg**

# Faculty of Engineering and Natural Sciences

# Department of IT

PROJECT PAPER

# **TWITTER DATA ANALYSIS USING NLP**

Supervisor

Prof. Dr. Dzelila Mehanovic

SARAJEVO

December 10, 2023

*Table of contests*

1. SUMMARY………………………………………………………
2. INTRODUCTION………………………………………………..
3. THEORETICAL BACKGROUND………………………………
4. LITERATURE REVIEW…………………………………………
5. RESEARCH QUESTION/HYPOTHESIS………………………..
6. METHODOLOGY………………………………………………..
7. DATA AND FINDINGS………………………………………….
8. CONCLUSION………………………………………………………
9. REFERENCES………………………………………………………
10. USEFUL SOFTWARES AND EXTENSIONS…………………..
11. CITATIONS AND REFERENCES………………………………

**SUMMARY**

The purpose of this research is to perform sentiment analysis on Twitter data using Natural Language Processing (NLP) techniques, particularly leveraging the NLTK library in Python within a Jupyter notebook environment. The study aims to explore sentiment classification methods, evaluating the emotional tone of tweets and categorizing them as neutral, positive, or negative sentiments, utilizing NLTK's SentimentIntensityAnalyzer.

The sample consists of Twitter data with columns like 'Tweet' and 'Sentiment' sourced from a CSV file. The methodology involves tokenizing and processing the text, grading sentiment, counting occurrences of the hashtag #fifa, and analyzing word frequencies. The research extends its analysis by employing wordclouds to visually represent the most common words and their prevalence in the dataset.

Furthermore, the study investigates the impact of removing stopwords and explores the list of eliminated stopwords. The expected results include gaining insights into prevalent sentiments on Twitter regarding a specified topic, frequency of the hashtag #fifa, and a comprehensive understanding of word usage, visually depicted through wordclouds.

Possible limitations include inherent subjectivity in sentiment analysis, potential variations in language use, reliance on hashtag frequency as an indicator of topic prevalence, and the effectiveness of stopwords removal, which may be context-dependent. The addition of wordcloud analysis enhances the visual representation of the most frequent words, providing a holistic perspective on the dataset.

**INTRODUCTION**

This research dives into Twitter's world, specifically tweets with #fifa, using NLTK in Python to decode sentiments. The main goal is to create word clouds, spotlighting the trendiest words and those linked to #fifa. We're also testing how well our sentiment guesses match real sentiments. Plus, we're poking at word variations using Lancaster and Porter stemmers.

Why word clouds? Well, they're like visual summaries, making the data fun and digestible. By doing this, we hope to uncover the vibe around #fifa on Twitter and see how sentiment analysis and word variations play out in the social media chatter. It's all about decoding the Twitter talk on #fifa in a snappy, visual, and insightful way! 🌐⚽✨

# **General notes**

The project paper should be written in academic writing style and submitted in typescript and free of any typographical errors. Please use Times New Roman, 12 pt., margins narrow, spacing 1.0 with headings in uppercase and bold and subheadings in italic. Number of pages cannot be less than 5 nor bigger than 7 pages (not including title page, list of references and appendices). Title page should be prepared according to the standard University guide.

# **Sections of Project Paper**

The Project paper should be comprised of the following1:

* 1. **Title** – Title of the Project
  2. **Summary** – Using 200 words, describe the purpose, theoretical background on which you are grounding your research, sample and methodology used, expected results and possible limitations.
  3. **Introduction** – Briefly explain the background of the research, clarify the major issues and the reasons why you decided to focus on them. Concisely present your research statement.
  4. **Theoretical Background** – Reference works that are necessary for the analysis and are crucial to your research. (Optional)
  5. **Literature review –** Describe, in detail, the development of the field of your interest in terms of research that has been conducted in your field of choice so far. Be sure to include most of the recent research.
  6. **Hypotheses (Research Questions) and Research Model –** Present the hypotheses/research questions you have developed based on the literature, as well as model you are proposing.
  7. **Methodology -** Present your sample design and data collection, research instrument(s), research material used, means of data analyses. Present these within the context of your research (in terms of data collection)
  8. **Data and findings(results) –** Describe the data you have collected and the findings of your research.
  9. **Conclusion –** In short, present the conclusions of your research.
  10. **References –** List all resources you have consulted during your writing.

# **Example of Project Paper**

INTERNATIONAL BURCH UNIVERSITY

FACULTY OF EDUCATION AND HUMANITIES

DEPARTMENT OF ELECTRICAL AND ELECTRONICS ENGINEERING

A picture containing text

Description automatically generated

Write the project title here (ALL CAPITAL 14 pt, not bold)

PROJECT PAPER

Supervisor

Prof. Dr. Add supervisor's Name and Surname

SARAJEVO

Choose month 2022

## **SUMMARY**

Because of the growing effects of the globalization in various business environments, the manufacturing industry is expected to be effective and yet efficient. According to this, in planning, scheduling and controlling a project, which is a combination of various activities, project management techniques (PERT and CPM) are used. Therefore, the research question is how will the implementation of CPM and PERT influence the effectiveness and efficiency of furniture company ''Dallas''? The answer to this question is relevant in order to point out the importance of those methods in reducing the project completion time and costs. The data are taken from the furniture company ''Dallas'' and it will be combined with literature reviews. The research study is fueled by the following objectives: First is to determine the activities that are involved in the manufacturing process in selected company. Second is to demonstrate the benefits, as well as the drawbacks that those methods might create in the organization. And third is to demonstrate the influence of CPM and PERT in the entire furniture industry and its competitiveness. Implications of this research paper are evaluation of the project completion time and control of the resources, in order to see that the project is completed within the planned time and cost by using mentioned methods. At the end of the study, the result is expected to help all the [individuals](http://ivythesis.typepad.com/term_paper_topics/2008/05/collaboration-a.html) as well as the [companies](http://ivythesis.typepad.com/term_paper_topics/human_resource_management_of_airline_companies/) to understand more the concept of CPM and PERT methods in reducing the project completion time and costs.

**Keywords:** CPM, PERT, Furniture Company, Optimization

## **INTRODUCTION**

Planning, Scheduling (or organizing) and Control are considered to be basic Managerial functions, and CPM/PERT has been rightfully accorded due importance in the literature on Operations Research and Quantitative Analysis. Far more than the technical benefits, it was found that PERT/CPM provided a focus around which managers could brain-storm and put their ideas together. Most important, it became a useful tool for evaluating the performance of individuals and teams. There are many variations of CPM/PERT which have been useful in planning costs, scheduling manpower and machine time.

The research study is fueled by the following objectives: First is to determine the activities that are involved in the manufacturing process in selected company. Second is to demonstrate the benefits, as well as the drawbacks that those methods might create in the organization. And third is to demonstrate the influence of CPM and PERT in the entire furniture industry and its competitiveness. This paper comprises the possibility to generate importance of CPM and PERT methods in reducing the project completion time and costs in furniture industry. The study can gain advantages that are helpful in the continuous progress of the investigation. One of those advantages is to cover the literature gaps concerning the past studies related to the same subject. In addition, through the collection of information, the study can emphasize the idea about the methods applied in production process in furniture industry.

**THEORETICAL BACKGROUND**

This study directly or indirectly deals with systems theory and TQM theory which are explained in more detail under following sections. Systems theory has been developed by Ludwig von Bertalanffy in his article “General Systems Theory: A New Approach to Unity of Science” in which he presented the foundations for new theory (Bertalanffy, 1951). The origins of TQM theory go back to the work of W. Edwards Deming. His practice of TQM in industry and development of TQM as a discipline of study is significant for scholars and practitioners (Rungtusanatham, et al., 2003).

## **LITERATURE REVIEW**

Since the development of CPM and PERT during the 1950s, the techniques have been the subject of hundreds of research papers, but little work has been done in the area of the time-cost problem in furniture industry. Research has generally been focused on PERT, since the deterministic CPM presents few problems of interest.[[1]](https://paperpile.com/c/EXGcvP/QXlP) demonstrated several advantages of applying simulation techniques to PERT, including more accurate estimates of the true project length, flexibility in selecting any distribution for activity times, and the ability to calculate "criticality indexes," which are the probability of various activities being on the critical path. Pritsker and Happ (1966) developed a modification of PERT called the Graphical Evaluation and Review Technique (GERT). GERT allowed activity times to follow several different distributions. Project completion time distributions were computed through Monte Carlo simulation.

Kennedy and Thrall (1976) developed a modification of GERT called Project Length Analysis and Evaluation Technique (PLANET). PLANET added the ability to calculate the probability of activities being critical and find the distribution of completion times for each activity. Ramini (1986) proposed an algorithm for crashing PERT networks with the use of criticality indices. Bottlenecks traditionally have multiple feeds into a very narrow path that is critical to the project's completion. Johnson and Schon (1990) used simulation to compare three rules for crashing stochastic networks-also involving the use of criticality indices. Feng, Liu, & Burns, (2000) presented a hybrid approach that combines simulation techniques with a genetic algorithm to solve the time-cost trade-off problem under uncertainty. Lu and Li (2003) suggest that redundant relationships be removed before the backward pass, but they do not provide a procedure as to how to remove them. Lu’s method may generate a large number of redundant resource links as demonstrated in the example because it does not consider original technological links of the CPM network when resource links are identified.

Additional authors which have studied various PERT problems via simulation include Klingel (1966), Gray (1969), Burt (1971), Herbert (1979), Schonberger (1981), and Dodin (1984), and

Kidd (1986).

## **RESEARCH QUESTION/ HYPOTHESIS**

The following is the basic question of research paper: *How will the implementation of CPM and PERT influence the effectiveness and efficiency of furniture company ''Dallas''?*

If there are numerous possible critical paths the probability may be much less than 50%. This may be costly if there are penalties for late completion of the project. The complete distribution of project completion time needs to be considered when crashing. Since there may be numerous possible critical paths, crashing a given activity by one time period will not necessarily reduce the completion time of the project by one time period. The expected reduction in project completion time must be considered in addition to the time/cost slope when selecting an activity to be crashed.

Hypothesis: *CPM and PERT considerably reduce the project completion time in furniture company Dallas.* At the end of the study, the result is expected to help all the [individuals](http://ivythesis.typepad.com/term_paper_topics/2008/05/collaboration-a.html) as well as the [companies](http://ivythesis.typepad.com/term_paper_topics/human_resource_management_of_airline_companies/) to understand more the concept of CPM and PERT methods in reducing the project completion time and costs. Apparently, it is expected that final, quantitative results will point out the importance of implementing those methods in planning, scheduling and controlling a project in terms of providing effectiveness and efficiency of furniture company. To test hypothesis, six products from different product lines have been selected. In order to define ‘critical path’, project duration and cumulative project costs for all of this products, time and cost of all activities in production process need to be presented in table. To do the research successfully, literature reviews related to this field in other industries will be used. Unfortunately, there is a lack of literature reviews of CPM and PERT in furniture industry, which presents a great challenge. The obtained findings are expected to show that these techniques considerably reduce the project completion time. All findings will be compared with previous data of the company (project duration and cumulative project costs of selected products). If they confirm the hypothesis, this paper will help to all managers in furniture industry to implement CPM and PERT to their projects, and by doing that, they will improve effectiveness and efficiency of their organizations. Furthermore, it may challenge other researchers to fulfil gaps in literature reviews related to this topic.

## **METHODOLOGY**

The applied method in the study is the processing of primary data as a result of direct interview with an expert team in furniture company. The data will be selected respecting the Time schedule (Table 1). In primary data collection, qualitative and quantitative methods can be used, as primarily quantitative study may uses qualitative results to help interpret or explain the quantitative findings. The key point here is that the collected data are unique to this research and, until it is publish, no one else has access to it. It is important to know in advance what questions to ask/areas to cover – they are ‘imposed’. When all needed data are provided, then the steps of CPM and PERT can be followed. Through this method, the study can gain advantages that are helpful in the continuous progress of the investigation. It is also important to discover the literature gaps because of the review done to the past studies concerning the same subject. In addition, through the collection of information, the study can emphasize the idea about the CPM and PERT applied in furniture company.

## **DATA AND FINDINGS**

## In this part the results of the data analysis are presented. The data were collected and then processed in response to the problems. Two fundamental goals drove the collection of the data and the subsequent data analysis. Those goals were to develop a base of knowledge about the technology education curriculum organizer, construction, as it is perceived and utilized relative to other curriculum organizers, and to determine if current perception and utilization are consistent with the basic goals or principles of technology education. These objectives were accomplished. The findings presented in this chapter demonstrate the potential for merging theory and practice.

## **CONCLUSION**

The aim of this study was to find correlation between…Sample was comprised from…Data was collected through the means of…Data were processed analyzed with usage of … method. Findings showed that…Therefore we can conclude that…

This study will bring benefit to the several key groups. Benefit for government is…On the other hand benefit for scholarly world and future research is…

## 

## **REFERENCES**

[1] [Federation of American Societies for Experimental Biology, *Chemistry in medicine: A Symposium held on the 80th birthday of Donald Dexter Van Slyke, Brookhaven National Laboratory, 1963*. Federation of American Societies for Experimental Biology, 1963.](http://paperpile.com/b/EXGcvP/QXlP)

[2] [A. Khanna, D. Gupta, Z. Pólkowski, S. Bhattacharyya, and O. Castillo, *Data Analytics and Management: Proceedings of ICDAM*. Springer Nature, 2021.](http://paperpile.com/b/EXGcvP/67dy)

# **Useful softwares and extensions:**

* Grammarly: <https://app.grammarly.com/> - used for grammar check and suggestions
* Paperpile: <https://paperpile.com/app> - used for citations and references
* QuillBot: <https://quillbot.com/> - used for paraphrasing and plagiarism check
* Google Scholar - used for finding useful research papers [[2]](https://paperpile.com/c/EXGcvP/67dy)

# **Citations and References:**

Each paper should have at least 10 citations, meaning you have to find similar research papers and cite them in your paper as a literature review. These citations can be in different formats like [APA](https://guides.libraries.psu.edu/apaquickguide/intext#:~:text=APA%20in%2Dtext%20citation%20style,2005%2C%20p.%2014).) or [IEEE](https://researchguides.njit.edu/ieee-citation/ieeereferencing).

For direct quotations, we will cite it “***as mentioned earlier (Field, 2005, p. 14) …*”**.

To add a reference in APA style we use this format:

*Derwing, T. M., Rossiter, M. J., & Munro, M. J. (2002). Teaching native speakers to listen to foreign-accented speech. Journal of Multilingual and Multicultural Development, 23(4), 245-259.*

IEEE citation style includes in-text citations, numbered in square brackets, which refer to the full citation listed in the reference list at the end of the paper. The reference list is organized numerically, not alphabetically.

For direct quotations, we will cite it “***as mentioned earlier [3], [4]­–[6], [8] …*”**

And to add reference in IEEE style we use this format:

*[1] W. Brown, "Electrical Design Considerations," in Advanced Electronic Packaging: With Emphasis on Multichip Modules: Wiley-IEEE Press, 2013, pp. 51-74.*

**PLEASE CREATE A PROJECT PAPER INSIDE THE** [**FOLDER**](https://drive.google.com/drive/u/1/folders/1FkctcIvMo1MjIK6l4UG3pVFi9Bb5iQZA) **FOR YOUR PROJECTS SO IT IS EASIER TO REVIEW THEM.**