|  |  |
| --- | --- |
|  | Hacettepe UniversityComputer Engineering DepartmentBBM479 Project Proposal Report |

**Project Details**

|  |  |  |
| --- | --- | --- |
| **Title** | A data driven tool for generating social SE insights into software development | |
| **Short Description**  **(max. 200 words)** | This project, designed for student exploration, focuses on investigating the connection between issue comments and the issue lifecycle in the context of software development. By utilizing data from GitHub repositories, we aim to gain insights into the social aspects of software engineering. This research project is structured as follows: | |
| **Supervisor** | **Dr. Tuğba GÜRGEN ERDOĞAN** | |
| **Technical and Scientific Difficulty** | ( ) Easy ( ) Mediocre ( ) Challenging | |
| **External Support** | ( ) Yes ( X ) No | |
| **If yes,** | **Type** | **Details** |
| ( ) Company Funding / Support | Company name:  Amount: |
| ( ) TÜBİTAK Project Fund | Type:  Amount: |
| ( ) Other Fund | Source :  Amount: |

**Group Members**

|  |  |  |
| --- | --- | --- |
|  | **Full Name** | **Student ID** |
| 1 | İbrahim Enes Genişyürek | 21892757 |
| 2 | Oğuzhan Taşçı | 2200356842 |
| 3 |  |  |
| 4 |  |  |

**Project Summary ( / 20 Points)**

Explain the project in summary, including your motivation to do the project, your solution plan in short and your expected outcome and impact. You have to summarize your project between 200-500 words.

|  |
| --- |
| Project Summary:  Our project, "Exploring Emotional Impact in Software Development," is driven by the need to understand how emotions expressed in issue comments on GitHub influence the software development process. Our motivation stems from the recognition that effective collaboration and timely issue resolution are vital in software development. Emotional content in communication can significantly impact team dynamics and project outcomes.  Motivation:  In the software development world, issues and their resolutions are the building blocks of success. How team members interact, communicate, and support each other emotionally can affect the efficiency and effectiveness of issue resolution. We are motivated to delve into this aspect to help developers, project managers, and the software development community create healthier and more productive environments.  Solution Plan:  Our solution plan involves comprehensive data analysis. We collect data from GitHub repositories, focusing on issues, issue comments, and relevant timestamps. After preprocessing the data, we apply sentiment analysis to assess the emotional tone of issue comments. Aligning timestamps with issue lifecycle events, we examine how emotions in comments relate to the different phases of an issue's lifecycle.  Expected Outcome:  We anticipate gaining valuable insights into the social dynamics of software development. We aim to determine whether specific emotional content in issue comments influences issue resolution times, community collaboration, and project success. These insights can have a profound impact on the software development community by improving team interactions, enhancing project management strategies, and fostering a more positive and productive environment.  Impact:  The impact of this project extends to various stakeholders within the software development ecosystem. For developers, it offers a better understanding of how their communication can affect project outcomes, promoting more empathetic and constructive interactions. Project managers gain insights that can guide team management and issue resolution strategies, leading to more efficient and successful projects. Researchers can leverage our findings to explore the social dimensions of software development further.  In summary, our project's motivation is to uncover the emotional undercurrents in software development communication. Our solution plan involves rigorous data analysis to explore the relationship between issue comments and the issue lifecycle. The expected outcome is a deeper understanding of how emotions impact software development, with potential impacts on team dynamics, project success, and the broader software development community. |

**Problem Definition and Literature Review ( / 20 Points)**

Define your problem as clearly as possible. Explain your inputs, your context, your outputs and your limitations. Try to use a scientific language as much as possible. Where necessary use citations to existing literature to create context and clarify the problem. Equations, flow charts, etc. are welcome.

|  |
| --- |
| Problem Statement:  The problem addressed in this research is the investigation of the relationship between emotional content within issue comments on GitHub and the various phases of the software development issue lifecycle. Specifically, we seek to discern how the emotional sentiment expressed in issue comments influences the efficiency and dynamics of issue resolution, collaboration, and the overall success of software development projects.  Context:  Software development projects hosted on GitHub are dynamic and collaborative environments where developers and team members interact through issue comments to identify, discuss, and resolve software-related problems. The issue lifecycle encompasses stages such as issue creation, assignment, discussion, coding, testing, and eventual closure. Emotions play an integral role in communication, impacting the interpersonal dynamics among developers and, consequently, the progress of issues.  Inputs:  The primary input for this research is the dataset extracted from GitHub repositories. It includes information on issues, issue comments, and relevant timestamps. Additionally, natural language processing tools are employed to analyze the emotional content of issue comments. Sentiment analysis models are applied to classify comments into positive, negative, or neutral sentiments. Timestamp alignment ensures that issue comments are linked to specific phases of the issue lifecycle.  Outputs:  The outputs of this research include insights, data visualizations, and statistical analysis. The key findings reveal the relationships between emotional content within issue comments and the issue lifecycle stages. Data visualizations, such as graphs and charts, illustrate these relationships, and statistical tests provide quantitative assessments of the observed patterns. The project's ultimate output is a comprehensive understanding of how emotions impact software development processes.  Limitations:  Data Quality: The quality and completeness of the dataset play a critical role in the research. Incomplete or inaccurate data may lead to biased results.  Emotion Detection Accuracy: The accuracy of sentiment analysis tools can influence the quality of emotional content classification in issue comments. The limitations of these tools should be considered.  Generalizability: Findings may be specific to the context of GitHub and may not be entirely generalizable to all software development platforms or organizations.  Causality: While this research can identify correlations, establishing causality between emotional content and issue resolution times may be challenging. The research is primarily observational in nature.  Privacy and Ethics: Ethical considerations, including the privacy of developers, should be taken into account when handling personal data in issue comments.  In addressing these challenges, the research strives to provide a comprehensive and meaningful exploration of the impact of emotions on software development practices, acknowledging the multifaceted nature of this complex issue. |

**Solution Plan ( / 20 Points)**

Explain the potential paths to solution. You should propose at least one solid plan to attack the problem. Dissect your plan into steps and clearly identify the inputs and outputs of each step. You are not expected to provide the technical details of each step. Provide a weekly timeline/Gantt chart displaying the relevant weeks for each step.

|  |
| --- |
| Weekly Timeline:  Weeks : Data Collection and Preprocessing  Weeks : Sentiment Analysis  Weeks : Temporal Alignment  Weeks : Data Analysis  Weeks : Data Visualization  Weeks : Interpretation and Insights  Weeks : Documentation and Reporting  Week : Presentation and Communication  Weeks : Feedback and Refinement  Step 1: Data Collection and Preprocessing (Week )  Input: GitHub repositories data containing issues, issue comments, and timestamps.  Output: A cleaned and well-structured dataset ready for analysis.  Step 2: Sentiment Analysis (Week )  Input: Cleaned dataset.  Output: Classification of issue comments into positive, negative, or neutral sentiments.  Step 3: Temporal Alignment (Week )  Input: Dataset with sentiment labels.  Output: Aligned timestamps of issue comments with issue lifecycle stages.  Step 4: Data Analysis (Week )  Input: Aligned dataset.  Output: Statistical findings and patterns regarding the emotional content's impact on the issue lifecycle.  Step 5: Data Visualization (Week )  Input: Statistical findings.  Output: Visual representations (graphs, charts) illustrating the relationships between emotional content and the issue lifecycle.  Step 6: Interpretation and Insights. (Week )  Input: Visualizations and statistical findings.  Output: Insights regarding the impact of emotions on software development processes.  Step 7: Documentation and Reporting (Week )  Input: Insights and findings.  Output: A comprehensive project report detailing the research process and outcomes.  Step 8: Presentation and Communication (Week )  Input: Project report.  Output: Presentation of findings to stakeholders, developers, and the research community.  Step 9: Feedback and Refinement (Week )  Input: Feedback from stakeholders.  Output: Iterative improvements to the research and project documentation. |

**Methodology ( / 20 Points)**

Explain the methodology you will use in each of the steps you have described under your solution plan. Here, you are expected to give more technical details about each solution step. Also explain how each member of the project will contribute by assigning members to steps. If you are assigning more than one member to a step, explain their specific role and how the work will be divided among them.

|  |
| --- |
| Methodology:  Data Collection: Gather data from GitHub repositories, specifically focusing on issues, issue comments, and their associated timestamps.  Data Preparation: Clean and preprocess the data, ensuring it is ready for analysis.  Sentiment Analysis: Apply sentiment analysis techniques to evaluate the emotional tone of issue comments.  Timestamp Alignment: Align timestamps of issue comments with key issue lifecycle events, such as issue creation and closure.  Data Analysis: Analyze the data to explore the relationship between emotional content in issue comments and the issue lifecycle.  Results Visualization: Create visual representations of the findings for better comprehension.  Step 1: Data Collection and Preprocessing:  Methodology: Use GitHub API or data scraping tools to collect issue-related data. Preprocess the data by removing duplicates, handling missing values, and ensuring data consistency.  Members: Member A will be responsible for data collection, while Member B will focus on data preprocessing.  Step 2: Sentiment Analysis:  Methodology: Employ pre-trained sentiment analysis models (e.g., VADER, TextBlob) or build a custom model using natural language processing techniques to classify issue comments into positive, negative, or neutral sentiments.  Members: Member A will lead the sentiment analysis task, and Member B will assist in model selection and evaluation.  Step 3: Temporal Alignment:  Methodology: Utilize timestamps to align issue comments with key issue lifecycle events (e.g., issue creation, closure). Ensure precise matching of comments with corresponding lifecycle phases.  Members: Member A will oversee the temporal alignment process, while Member B will verify the accuracy of the alignment.  Step 4: Data Analysis:  Methodology: Employ statistical techniques, such as regression analysis or hypothesis testing, to analyze the relationships between emotional content in issue comments and the issue lifecycle. Evaluate if certain sentiments impact resolution times.  Members: Member A will conduct the statistical analysis, while Member B will assist in data interpretation.  Step 5: Data Visualization:  Methodology: Create visual representations (charts, graphs) to illustrate the relationships discovered in the data analysis. Use data visualization tools such as Matplotlib or Tableau.  Members: Member A will generate visualizations, while Member B will help design the visual elements and interpret the visualizations.  Step 6: Interpretation and Insights:  Methodology: Analyze the visualizations and statistical findings to draw meaningful insights about how emotions impact the issue lifecycle. Consider the implications for software development.  Members: Member A will lead the interpretation, while Member B will critically review and validate the insights.  Step 7: Documentation and Reporting:  Methodology: Prepare a comprehensive project report that documents the entire research process, findings, and insights. Ensure clarity and readability.  Members: Member A will compile the report, while Member B will review and edit for clarity and coherence.  Step 8: Presentation and Communication:  Methodology: Create a presentation summarizing the research for stakeholders and the research community. Use visual aids to convey key insights effectively.  Members: Member A will design the presentation, while Member B will provide assistance in delivering it effectively.  Step 9: Feedback and Refinement:  Methodology: Collect feedback from stakeholders and users regarding the project and research findings. Evaluate the suggestions for potential refinements.  Members: Both Member A and Member B will collaborate in gathering and analyzing feedback and jointly decide on refinements. |

**Outcome and Impact ( / 20 Points)**

Explain the expected outcome of your project. If it is a software product, try to include example screen designs, if it is a hardware product, try to provide detailed technical specifications, if it is research output try to explain the outcome’s contribution to the field. Also, explain the potential impacts of your results. These may be how the result will be used in real life, how it will change an existing process, or where it will be published, etc.

|  |
| --- |
| Expected Outcome:  The expected outcome of our project, "Exploring Emotional Impact in Software Development," is multi-faceted and extends to both research findings and practical implications:  Research Findings:  Relationships Between Emotions and Issue Lifecycle: The project aims to provide clear insights into how emotional content in issue comments on GitHub is related to the various phases of the issue lifecycle. It will reveal patterns and correlations that showcase how emotions impact issue resolution times, collaboration dynamics, and project success.  Visual Representations: We anticipate generating visualizations, including charts and graphs, that visually depict the discovered relationships. These visual aids will help stakeholders and researchers easily grasp the emotional impact on software development processes.  Practical Implications:  Improved Project Management: The research findings can be applied to enhance project management strategies. By understanding the influence of emotions, project managers can make informed decisions to optimize team collaboration and issue resolution processes.  Enhanced Developer Interaction: Developers can utilize the insights to foster more empathetic and constructive interactions within the development community. Recognizing the impact of their words on issue outcomes can lead to more effective communication.  Research Community Contribution: The research results will contribute to the wider research community by shedding light on the social dimensions of software development. It may serve as a basis for further studies in the field of software engineering and human-computer interaction.  Impact:  The potential impacts of our results encompass several areas:  Real-Life Application: Software development teams can immediately apply the insights to improve their processes, leading to more efficient issue resolution and project success.  Process Improvement: The findings can lead to changes in existing software development processes by emphasizing the importance of managing emotions in online communication.  Publication: The research results are expected to be disseminated through academic publications, contributing to the scholarly understanding of emotional dynamics in software development.  Community Building: The insights can contribute to a more positive and collaborative software development community, fostering better relationships among developers.  In summary, the project's expected outcomes include valuable research findings and practical applications. The insights gained from this research have the potential to positively impact the software development community, guiding project management practices and fostering more effective developer interactions. The dissemination of these findings in academic and industry publications will further extend the reach and influence of the project's outcomes. |