

Hello, my name is Pawan Solanki, and I would like to share a brief description of my recent project.

1. **Project Title** - Natural Language Processing with Disaster Tweets

2. **About the project** –

It is a Kaggle competition and we are tasked with predicting whether a tweet is related to a disaster or not. Our objective is to analyse a dataset of labelled tweets and develop a predictive model that classifies each tweet accordingly. This involves preprocessing the text data, extracting relevant features, and applying machine learning techniques to train and evaluate our model. The ultimate goal is to accurately identify disaster-related tweets, which can enhance our understanding and response to such events.

3. **Technology Stack** –

In this project, I employed **Python** as the primary programming language to perform data analysis and manipulation, utilizing libraries such as Pandas and NumPy for efficient data handling. For visualization, I used Matplotlib to create informative charts and graphs. To preprocess the data and make predictions, I turned to **Scikit-learn**, which provided robust machine learning tools. Additionally, I integrated **NLTK** and **spaCy** for advanced natural language processing tasks, enabling deeper insights from textual data throughout the project.

4. **Challenges and Solutions** –

When I started working on the project, I encountered textual data that required preprocessing to remove unnecessary characters and words. To tackle this, I explored NLTK and SpaCy for effective text handling. I experimented with multiple models, including logistic regression and random forests, to identify the best performer. However, my initial model suffered from overfitting, so I focused on feature extraction techniques to mitigate this issue and improve performance.

5. Team Size –

I worked on this project as individually.

6. Project Outcome –

After completing this project, I achieved a reliable predictive model for classifying disaster-related tweets, showcasing improved accuracy and relevant performance metrics. I enhanced my technical skills in text preprocessing, feature extraction, and the use of tools like NLTK and SpaCy, while gaining a deeper understanding of the challenges in natural language processing. Additionally, this project strengthens my portfolio, highlighting my capabilities in data science and potentially opening networking opportunities within the field.

7. Link of my Notebook – [Click Here](#)