**Own project proposal**

**Basic Info:**

**Project Title: Creating a Sentiment Analysis Framework using Spark and Machine Learning techniques to analyze large amounts of Twitter data related to the Ukraine Conflict.**

**Background and Motivation:**

Social media sites, especially Twitter, have seen a lot of conversation about the current violence in Ukraine. The sentiment analysis of tweets on the conflict can offer important insights into how the public feels about it. I'm inspired to learn more about this subject since I'm a data science enthusiast with a passion for social media analysis.

**Project Objectives:**

The primary questions for this project aims to answer are:

1.What are the main topics or themes discussed in tweets related to Ukraine, and how can they provide insights into public perception of the conflict?

2.How do sentiment scores in tweets related to the Ukraine conflict vary over time and geography on social media platforms, and can the insights gained from this analysis be used to identify patterns or trends in public opinion?

3.Can machine learning algorithms like Spark accurately classify the sentiment conveyed in tweets about the Ukraine conflict?

4.How can the developed sentiment analysis system be further improved to enhance its accuracy and efficiency? What is the significance of comparing the accuracy of different sentiment analysis techniques in the context of the Ukraine conflict, and what knowledge can be gained from this analysis?

**Data:**

The data for this project will be collected from the "🇺🇦 Ukraine Conflict Twitter Dataset" by BwandoWando available on Kaggle.com. The dataset contains tweets monitoring the ongoing Ukraine-Russia conflict.

**Ethical considerations:**

This project involves the sentiment analysis of tweets related to sensitive topics such as war and conflict, which requires ethical considerations to ensure the integrity of the analysis. This involves protecting user privacy and anonymity, avoiding misleading information, and acknowledging any biases or limitations in the dataset and analysis. The research will follow principles of informed consent, data protection, and transparency in reporting the results.

**Data Processing:**

The data will be preprocessed by cleaning and filtering out irrelevant tweets, removing stop words, and applying stemming or lemmatization techniques. The sentiment analysis will be performed using Spark and machine learning techniques such as Naive Bayes and Support Vector Machines (SVM) to classify the sentiment of each tweet as positive, negative, or neutral.

**Exploratory Analysis:**

The exploratory analysis will involve visualizations such as word clouds, bar charts, and line graphs to identify the main topics or themes discussed in tweets related to Ukraine, and how they provide insights into public perception of the conflict.

**Analysis Methodology:**

The sentiment analysis system will be developed and evaluated using metrics such as precision, recall, and F1 score. The system will be improved by incorporating advanced machine learning algorithms. The sentiment scores of the tweets related to the Ukraine conflict will be analyzed over time and geography, and patterns or trends in public opinion will be identified.

**Interpretation:** Interpreting the results to gain insights into public perception of the Ukraine conflict and its impact on social media.

**Project Schedule:**

Week 1-2: Data collection and preprocessing

Week 3-4: Exploratory analysis and visualization

Week 5-8: Sentiment analysis using Spark and machine learning techniques

Week 9-10: System evaluation and improvement

Week 11-12: Results interpretation and report writing

**References**

Sentiment Analysis, Tweet Analysis and Visualization on Big Data Using Apache Spark and Hadoop <https://findit.southwales.ac.uk/permalink/44WHELF_USW/th9aii/cdi_proquest_journals_2512962633>

Public Sentiment and Emotion Analyses of Twitter Data on the 2022 Russian Invasion of Ukraine [**https://www.researchgate.net/publication/361218371\_Public\_Sentiment\_and\_Emotion\_Analyses\_of\_Twitter\_Data\_on\_the\_2022\_Russian\_Invasion\_of\_Ukraine**](https://www.researchgate.net/publication/361218371_Public_Sentiment_and_Emotion_Analyses_of_Twitter_Data_on_the_2022_Russian_Invasion_of_Ukraine)