"An Unsupervised Fuzzy Rule-Based System for Emotion Detection"

- **Research Areas:** XAI, NLP, ML, NNs, Fuzzy Systems
- **Programing Language:** Python
- **Libraries:** NumPy, Pandas, NLTK, Scikit-Fuzzy, Scikit-Learn, Keras, Tensorflow, Matplotlib
- **Models:** Trapezoidal Fuzzy System, Naïve Bayes, SVM, LSTM

The proposed fuzzy system:

I explored sentiment extraction from tweets using fuzzy logic. Unlike data-driven models such as **LSTM** and **CNN**, my approach utilizes a rule-based system with nine rules, trapezoidal membership functions, and SOM/Centroid for the defuzzification stage. I comprehensively compared the proposed model against other methods, including **Naïve Bayes** and **SVM using TF-IDF Vectorizer**, across five datasets from Kaggle and GitHub. Moreover, I compared the proposed fuzzy system with two other articles that used CNN and the triangular fuzzy model for sentiment analysis. The reason I chose these methods was to give a thorough comparison between black-and-white box models and conclude that our approach, despite being relatively simple, is interpretable, can outperform other models most of the time, and is best when dealing with the ambiguity of human language.

Datasets:

The largest dataset was **sentiment140** from Kaggle containing sentiments of 1.6 million tweets in 3 classes the second largest from GitHub containing around 27000 tweets from users. Overall, I tested the model across five different datasets including users' daily tweets, Apple product reviews, and nuclear tweets.

Results:

the results demonstrated superior performance in three of the five datasets for intricate models, such as SVM, and complete superiority in trapezoidal membership functions over triangular ones in fuzzy systems. Moreover, our method can be used in fields such as law enforcement and medicine where gathering data can be a challenging task. Furthermore, I incorporated **Reinforcement Learning** with fuzzy logic to optimize lexicon selection for Twitter sentiment analysis, finding that **Vader** and **AFINN** lexicons yielded the best results.

Achievements:

In conclusion, I was able to learn various NLP techniques from the research experience, such as **preprocessing methods (text cleaning, tokenization, lemmatization, …), various Lexicons, Python NLTK, and other NLP libraries, Tensorflow, PyTorch (for DNNs), Metric Libraries, RL**, and **most importantly, researching** as I had to read/analyze articles and make comparisons when relevant. Finally, the experience inspired me to look for other state-of-the-art approaches such as attention models and statistical models such as generalized Bayesian models to combine them with fuzzy logic.