Urdu Text Classification System (A Hands on workshop)

Description

Text classification is the process of assigning tags or categories to text according to its content. It's one of the fundamental tasks in Natural Language Processing (NLP) with broad applications such assentiment analysis, topic labeling, spam detection, and intent detection.

We will pick an annotated dataset of English, Roman Urdu and Urdu for text classification and create different sets of features. We will use different features to train different machine learning model and show its impact on accuracies.

Aims and Learning Outcome

The participants/audience will learn the following:

- What is text classification?
- How does text classification work?
- What are the most common business applications?
- What are the algorithms used for classifying text?

Outline

1.	Over view of python	60 min
2.	Machine Learning approaches	10 min
3.	Text Classification	10 min
4.	Introduction to sklearn	30 min
5.	Model development and testing	40 min

Profile of the Presenters

Dr. Muhammad Kamran Malik received the Ph.D. degree in computer science. He is currently an Assistant Professor with the Punjab University College of Information Technology (PUCIT), University of the Punjab, Lahore, Pakistan, in 2018. He has more than 17 years of teaching and development experience. He has authored 1 US patent and 30 journal and conference papers. His research interests include natural language processing, machine learning, and data science. He has provided Consultancy to many multinational firms on the above mentioned areas.

https://pucit.edu.pk/

https://www.ds-initiative.org/

Dr. Zubair Nawaz is an Assistant Professor at PUCIT. He holds a PhD in Computer Engineering from Delft University of Technology, 2011. His PhD thesis was related with the compiler optimizations for the Reconfigurable Computing (FPGAs). Currently, he is working on scaling data science applications to many-core processors for high efficiency. Broadly, his research interests are in the field of text mining, high performance computing, optimization and scientific computing. Dr. Nawaz has won Education Grant for CUDA Teaching Center Program in PUCIT. Grant includes 1 Tesla k40 and 5 Titan black GPUs having more than 15000 cores each.

https://pucit.edu.pk/

https://www.ds-initiative.org/