International Burch University
Faculty of Engineering and Natural Sciences
Department of Computer Science
CEN 569: Big Data Technologies

Project Proposal

10th November 2023

Group Members:

Semir Salkić
Emir Ribić

Topic Definition:

“Advancing Social Network Analysis: A Hadoop and Spark-Driven Approach to Link Prediction in Facebook100”

Short abstract:

Facebook provides valuable data for social network research, with link prediction techniques playing a key role. Our paper explores link prediction on the Facebook100 network, using Hadoop and Apache Spark for data processing.
Using Hadoop's distributed storage, we manage the Facebook100 dataset, ensuring scalability and fault tolerance. We employ Apache Spark for its fast in-memory data processing, improving the performance of machine learning algorithms.
We use Spark's analytics capabilities, including its MLlib for machine learning, to evaluate algorithms on the Facebook100 network. The paper outlines the methodology, from data preprocessing in Hadoop to model training and evaluation in Spark, comparing the performance of local and cloud-based approaches.
The use of Hadoop and Spark highlights the benefits of big data technologies for handling large-scale social network data.

Dataset details available at following LINKS (paper, dataset).