

## Master of Science Topics 2021-2022, AI-MAS Laboratory

### Machine Learning for NLP and Time Series Classification / Prediction Tasks

#### Title: Deep Detection of Fraudulent Transactions

Coordinator: SL.dr.ing. Dan Novischi([dan\\_marius.novischi@upb.ro](mailto:dan_marius.novischi@upb.ro))

#### Description:

Credit card fraud resulted in the loss of €1.80 billion in the EU and a soaring \$3 billion in the US to financial institutions between 2017 and 2018 alone. The rise of digital payments systems such as Apple Pay, Android Pay, and Venmo has meant that loss due to fraudulent activity is expected to increase. Moreover, fraud prevention systems are insufficient to provide adequate security to the electronic commerce systems. However, the collaboration between detection and prevention systems might be effective to secure electronic commerce and keep losses to a minimum. Nevertheless, there are many challenges that hinder the performance of fraudulent transaction detection, such as concept drift, real time detection, skewed distribution, large amounts of data and so on. Deep Learning presents a promising solution to this problem by enabling institutions to make optimal use of their historic customer data as well as real-time transaction details that are recorded at the time of the transaction.

The goal of this research project is to both develop an augmented balanced dataset of fraudulent transactions based on generative adversarial models, attention models and adaptive deep solutions for the detection of fraudulent transactions.

- [1] Ba, H. (2019). Improving Detection of Credit Card Fraudulent Transactions using Generative Adversarial Networks. arXiv preprint arXiv:1907.03355.7
- [2] Zeager, M. F., Sridhar, A., Fogal, N., Adams, S., Brown, D. E., & Beling, P. A. (2017, April). Adversarial learning in credit card fraud detection. In 2017 Systems and Information Engineering Design Symposium (SIEDS) (pp. 112-116). IEEE.
- [3] Carneiro, N., Figueira, G., & Costa, M. (2017). A data mining based system for credit-card fraud detection in e-tail. *DecisionSupport Systems*, 95, 91-101.
- [4] Abdallah, A., Maarof, M. A., & Zainal, A. (2016). Fraud detection system: A survey. *Journal of Network and Computer Applications*, 68, 90-113.

#### Title: Semantic Change Detection

Coordinators: S.L. Dr. Ing. Ciprian-Octavian Truică ([ciprian.truica@upb.ro](mailto:ciprian.truica@upb.ro))  
Conf. Dr. Ing. Elena-Simona Apostol ([elena.apostol@upb.ro](mailto:elena.apostol@upb.ro))

#### Description:

Semantic Change deals with the evolution of language over time. The main objective is to determine any changes that appear in historical data in the meaning of words. This research topic aims to design new Natural Language Processing, Machine Learning, and

Deep Learning techniques based on word and transformer embeddings in order to determine accurately diachronic information.

**Implementation Language:** Python

**Libraries:** Scikit-learn, PyTorch, Keras with Tensorflow, etc.

**Storage & Distribution Technologies:** MongoDB, Spark.

**No. Students:** 1

### **Title: Enhancing Scientific Collaborations**

*Coordinators:* S.L. Dr. Ing. Ciprian-Octavian Truică (ciprian.truica@upb.ro)

Conf. Dr. Ing. Elena-Simona Apostol (elena.apostol@upb.ro)

#### *Description:*

Scientific output has grown exponentially in the last years. As the number of papers has risen, there is a need to design new algorithms that integrate both graph structure and textual content to automatically analyze the scientific output and propose new work collaborations. This topic aims to design new Natural Language Processing, Machine Learning, and Deep Learning techniques for analyzing the scientific papers' textual content and Graph Mining techniques to propose new collaborations for scientists.

**Implementation Language:** Python

**Libraries:** Scikit-learn, PyTorch, Keras with Tensorflow, etc.

**Storage & Distribution Technologies:** MongoDB, Spark.

**No. Students:** 1

### **Title: Network-aware harmful content mitigation on social media**

*Coordinators:* S.L. Dr. Ing. Ciprian-Octavian Truică (ciprian.truica@upb.ro)

Conf. Dr. Ing. Elena-Simona Apostol (elena.apostol@upb.ro)

#### *Description:*

Harmful content consists of different types of textual data (e.g., hate speech, misinformation, fake news) that causes a person distress or harm, having a negative impact on the individual mental health, with even more detrimental effects on the psychology of kids and teenagers. This topic aims to design and implement new Network Analysis and Graph Mining models and strategies for mitigating the spread of harmful content on social media using information extracted from the diffusion medium.

**Implementation Language:** Python

**Libraries:** Scikit-learn, PyTorch, Keras with Tensorflow, etc.

**Storage & Distribution Technologies:** MongoDB, Spark.

**No. Students:** 1

### **Title: Network-aware fake news mitigation on social media**

**Coordinators:** S.L. Dr. Ing. Ciprian-Octavian Truică ([ciprian.truica@upb.ro](mailto:ciprian.truica@upb.ro))  
Conf. Dr. Ing. Elena-Simona Apostol ([elena.apostol@upb.ro](mailto:elena.apostol@upb.ro))

#### *Description:*

Fake news consists of news that is intentionally and verifiably false, and which could mislead readers by presenting alleged, imaginary facts about social, economic, and political subjects of interest. This topic aims to design and implement new Network Analysis and Graph Mining models and strategies for mitigating the spread of fake news on social media using information extracted from the diffusion medium.

**Implementation Language:** Python

**Libraries:** Scikit-learn, PyTorch, Keras with Tensorflow, etc.

**Storage & Distribution Technologies:** MongoDB, Spark.

**No. Students: 1**

### **Title: Context and network-aware fake news detection and mitigation**

**Coordinators:** S.L. Dr. Ing. Ciprian-Octavian Truică ([ciprian.truica@upb.ro](mailto:ciprian.truica@upb.ro))  
Conf. Dr. Ing. Elena-Simona Apostol ([elena.apostol@upb.ro](mailto:elena.apostol@upb.ro))

#### *Description:*

Fake news consists of news that is intentionally and verifiably false, and which could mislead readers by presenting alleged, imaginary facts about social, economic and political subjects of interest. This topic aims to design new Natural Language Processing, Machine Learning, Deep Learning, Network Analysis, and Graph Mining models, techniques, and strategies that use context and network information for mitigating the spread of fake news on social networks.

**Implementation Language:** Python

**Libraries:** Scikit-learn, PyTorch, Keras with Tensorflow, etc.

**Storage & Distribution Technologies:** MongoDB, Spark.

**No. Students: 1**

### **Title: Aspect-based Sentiment Analysis for Online Reviews**

**Coordinators:** S.L. Dr. Ing. Ciprian-Octavian Truică ([ciprian.truica@upb.ro](mailto:ciprian.truica@upb.ro))  
Conf. Dr. Ing. Elena-Simona Apostol ([elena.apostol@upb.ro](mailto:elena.apostol@upb.ro))

#### *Description:*

Aspect-based sentiment analysis (ABSA) is the task of identifying coarse-grained or fine-grained opinion polarity towards a specific aspect associated with a given target. Thus, the main objective of ABSA is to identify the entity E and attribute A pairs towards which an opinion is expressed in each sentence and determine the opinion's polarity. This research topic aims to design new Natural Language Processing, Machine Learning, and Deep Learning techniques in order to determine accurately the fine-grained opinion polarity from online reviews.

**Implementation Language:** Python

**Libraries:** Scikit-learn, PyTorch, Keras with Tensorflow, etc.

**Storage & Distribution Technologies:** MongoDB, Spark.  
**No. Students:** 1

### **Title: Text Simplification**

**Coordinators:** S.L. Dr. Ing. Ciprian-Octavian Truică ([ciprian.truica@upb.ro](mailto:ciprian.truica@upb.ro))  
Conf. Dr. Ing. Elena-Simona Apostol ([elena.apostol@upb.ro](mailto:elena.apostol@upb.ro))

#### *Description:*

Text Simplification is the process of reducing the linguistic complexity of a text, while retaining the original information content and meaning. This topic aims to design new Natural Language Processing, Machine Learning, and Deep Learning models and techniques that successfully manage to simplify textual content.

**Implementation Language:** Python

**Libraries:** Scikit-learn, PyTorch, Keras with Tensorflow, etc.

**Storage & Distribution Technologies:** MongoDB, Spark.

**No. Students:** 1

### **Title: Change Point detection for Time Series**

**Coordinators:** S.L. Dr. Ing. Ciprian-Octavian Truică ([ciprian.truica@upb.ro](mailto:ciprian.truica@upb.ro))  
Conf. Dr. Ing. Elena-Simona Apostol ([elena.apostol@upb.ro](mailto:elena.apostol@upb.ro))

#### *Description:*

The problem of change point detection in Time Series data deals with finding the point in time when the properties (e.g., mean, variance, etc.) of the Time Series change abruptly. Thus, a change point is a transition point between different states or continuous segments in the Time Series data. The aim of this research topic is to develop new Change Point Detection models and techniques using Statistics, Machine Learning and Deep Learning.

**Implementation Language:** Python

**Libraries:** Scikit-learn, PyTorch, Keras with Tensorflow, etc.

**Storage & Distribution Technologies:** MongoDB, Spark.

**No. Students:** 1

### **Title: Scalable Graph Mining Algorithms for Social Media Analysis**

**Coordinators:** S.L. Dr. Ing. Ciprian-Octavian Truică ([ciprian.truica@upb.ro](mailto:ciprian.truica@upb.ro))  
Conf. Dr. Ing. Elena-Simona Apostol ([elena.apostol@upb.ro](mailto:elena.apostol@upb.ro))

#### *Description:*

Graph Mining has seen an emerging use in Social Media Analysis with a focus on network analysis and information diffusion. As the graph dimensions are rapidly increasing current algorithms and strategies fall short when dealing in a timely manner with large data sets. Thus, this research topic aims to utilize Big Data Analysis technologies and Deep

Learning to improve Graph Mining algorithms by proposing new frameworks for graph processing and new algorithms and graph embeddings for large graph data analytics.

**Implementation Language:** Python

**Libraries:** Graphframe, PyTorch, Keras with Tensorflow, etc.

**Storage & Distribution Technologies:** Hadoop Ecosystem and Spark Environment.

**No. Students:** 1

### **Title: Anomaly detection for Time Series**

*Coordinators:* S.L. Dr. Ing. Ciprian-Octavian Truică ([ciprian.truica@upb.ro](mailto:ciprian.truica@upb.ro))

Conf. Dr. Ing. Elena-Simona Apostol ([elena.apostol@upb.ro](mailto:elena.apostol@upb.ro))

#### *Description:*

An outlier or an anomaly is a data point that significantly differs from other observations in a Time Series. Outliers can appear due to an experimental error or an anomaly in the measurement. Such suspicious points in the Time Series data must be identified and interpreted separately in order not to interfere with the analysis step and lead to wrong conclusions. The aim of this research topic is to develop new Anomaly Detection models and techniques using Statistics, Machine Learning and Deep Learning.

**Implementation Language:** Python

**Libraries:** Scikit-learn, PyTorch, Keras with Tensorflow, etc.

**Storage & Distribution Technologies:** MongoDB, Spark.

**No. Students:** 1