Shahid Beheshti University

# Machine Learning Fundamentals

## Final project– June 27, 2024

Hello everyone, I trust you are all in good health and spirits. This is the final assignment for our Machine Learning course. This document outlines three distinct projects that you can choose to undertake. Each project presents a unique challenge and opportunity to apply your ML skills to real-world problems. Please note that the deadline for this project is **July 10th**. Should you have any questions concerning the exercises, please feel free to reach out.

# Problem descriptions

## 1st problem

A fintech startup has gathered a dataset of crypto news articles and is eager to discover how these narratives influence Bitcoin's price. They require an analysis that correlates news sentiment with Bitcoin's price movements. Your task is to locate a dataset of Bitcoin prices corresponding to the news timestamps and conduct an exploratory data analysis. The end goal is to develop a predictive model that, by leveraging news attributes, can forecast whether Bitcoin's price will increase, remain stable, or decrease.

## 2nd problem

A bank requires a model for fraud detection in its transactions. Since there is no existing data collection process, you're tasked with identifying a robust dataset, conducting exploratory data analysis, and developing a model that can accurately flag fraudulent transactions. The selection of the dataset will be a critical component of your project and will be assessed accordingly.

## 3rd problem

A multinational corporation operates representative offices in 37 cities and requires hourly weather predictions to optimize their daily operations. The company seeks to predict hourly weather conditions to better prepare for daily activities. Utilizing historical hourly weather data, your goal is to construct a model that forecasts the weather for each city within 2012 to 2017. To achieve this, consider leveraging comparative analysis by using attributes and weather patterns from other cities as a comparative baseline to improve the accuracy of your forecasts. Incorporate regional climate trends and historical weather events that may influence the weather in similar cities. Utilize geospatial data to understand the geographical influences on weather patterns across different cities.

---

As you prepare to write your report on the process of solving each project, it's essential to follow a structured approach. This section outlines the critical steps you should incorporate into your report to demonstrate a clear and logical progression of your problem-solving journey.

## Problem Formulation

- **Clarifying questions**: What specific aspects of the problem do we need to address?
- **Assumptions**: What assumptions are we making about the data and the problem?
- **Translate an abstract problem into an ML problem**: How can we frame this as a machine learning task?
- **Do we need ML to solve this problem?** If so, why?

## Metrics Selection

- Explain the metrics you plan to use and justify why they are suitable for evaluating your model's performance.

## Exploratory Data Analysis (EDA)

- Conduct EDA to understand the data's structure, identify outliers, and detect patterns.
- Document your observations and insights in a way that informs subsequent steps.

## Feature Engineering

- Describe the feature engineering steps you take and explain why they are beneficial for your model.

## Model Evaluation

- Test different models or combinations of models and evaluate them to determine the best one.
- Include the results and evaluation process in your report.

## Demonstration App

- Building a demo app using frameworks like Gradio, … could earn you extra points.

Remember, thorough documentation and clear communication of your methodology and findings are crucial components of this assignment.

**Plagiarism will not be tolerated**. Homework submissions will be cross-checked against other students' submissions. Additionally, **the use of AI to fully generate answers or code** for assignments is strictly forbidden.