Class Project Proposal – Transformer Model for Bitcoin Price Prediction

### Course: Data 612 Deep Learning

### Team Members: Sirui Zeng,

### Date: June 27, 2025

### **1. Background and Significance**

Transformer models have revolutionized sequence modeling, particularly in NLP tasks, due to their ability to handle long-range dependencies using self-attention mechanisms. Recently, their success has extended into time series forecasting, offering advantages over traditional models like ARIMA and RNNs/LSTMs, which struggle with vanishing gradients or limited receptive fields.

In line with the course emphasis on modern neural network architectures (Lecture 4), we aim to apply a Transformer-based architecture to the problem of Bitcoin price prediction. This domain is well-known for its volatility and long-term dependencies, making it an ideal test case for self-attention models. Our objective is to demonstrate how Transformers can improve performance in forecasting complex, non-linear time series data such as cryptocurrency prices.

### **2. Statement of the Problem**

**Research Question:** Can a Transformer-based neural network accurately forecast short- to medium-term Bitcoin closing prices using historical OHLCV (Open, High, Low, Close, Volume) data?

**Sub-objectives:**

* Evaluate the Transformer's performance on multiple timeframes (15-minute, 1-hour, 4-hour, daily)
* Compare results against baseline models such as LSTM and linear regression
* Interpret the model through attention-based visualization

### **3. Dataset and Data Preparation**

We will use the publicly available **Bitcoin Historical Dataset (2018–2024)** from Kaggle:  
<https://www.kaggle.com/datasets/novandraanugrah/bitcoin-historical-datasets-2018-2024>

This dataset provides historical candlestick data (OHLCV) for Bitcoin (BTC/USDT), sourced from the Binance exchange via API, covering the period from January 1, 2018, to the present. It includes data in four timeframes: 15-minute, 1-hour, 4-hour, and daily. Each record includes the following fields: open time, open, high, low, close, volume, close time, quote asset volume, number of trades, taker buy volumes, and a placeholder field (ignore). The dataset is automatically updated daily through a custom script.