Overview

Volatility is one of the most prominent terms you'll hear on any trading floor – and for good reason. In financial markets, volatility captures the amount of fluctuation in prices. High volatility is associated to periods of market turbulence and to large price swings, while low volatility describes more calm and quiet markets. For trading firms like Optiver, accurately predicting volatility is essential for the trading of options, whose price is directly related to the volatility of the underlying product and in this project the goal is to accurately predict volatility. We will predict **the realized volatility of the next ten-minutes time window** with two data sets of the last ten minutes (600 seconds). One dataset contains ask and bid prices of almost each second, which allows us to calculate the realized volatility of the last ten minutes. The other dataset contains the actual record of stock trading, which is more sparse.

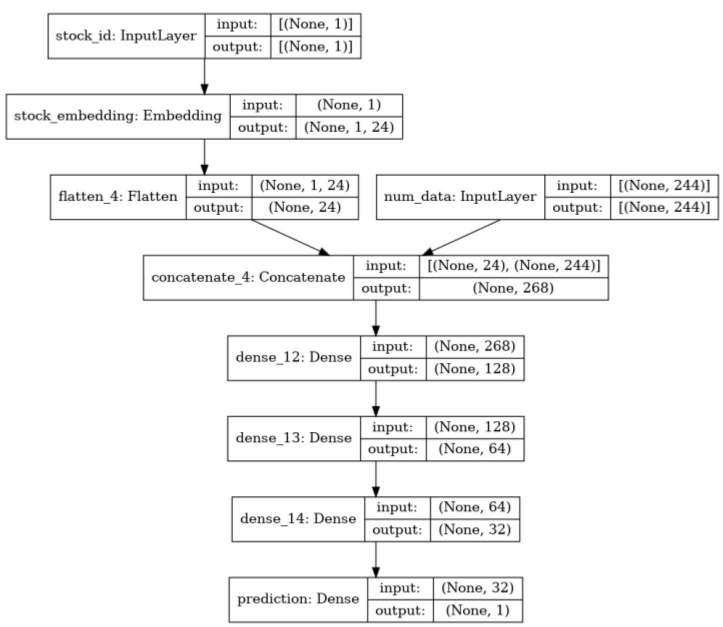
My approach for predicting volatility in 10 minute-windows consists of ensemble of 3 single models:

1. Tabnet: TabNet uses sequential attention to choose which feature to reason from at each decision step. It's sparse feature selection enables interpretability and better learning as the capacity is used for the most salient features and employs soft feature selection with controllable sparsity in end-to-end learning.

2. LightGBM: LightGBM is a gradient boosting framework that uses tree-based learning algorithms. It is designed to be distributed and efficient with the following advantages:

   - Faster training speed and higher efficiency
   - Lower memory usage
   - Better accuracy
   - Support of parallel, distributed, and GPU learning
   - Capable of handling large-scale data

3. Feedforward neural network (FFNN) : The neural network architecture which I used as one of the single models is as follows:



And the outcome is generated by weighted average of single models' outputs.

The table below lists the performance of the single models and the final blending in the scheme

		Local CV
Single model	Tabnet	0.24212
	LightGBM	0.23603
	FFNN	0.24817
Blend	(Tabnet+ LightGBM+ FFNN) *1/3	0.22010

The flowchart below outlines the technical details: