**Outline for Course Presentation**

**Capturing the Baijiu(白酒) Industry Landscape: A Textual Analysis Snapshot and Predictive Modeling of Stock**

**1. Introduction**

Analytical methods in digital economy can play a crucial role in financial forecasting, with a particular emphasis on stock prediction. The relevance of the digital economy to stock prediction is multifaceted. These methods involve the use of statistical techniques, machine learning algorithms, and other quantitative tools to analyze historical data and identify patterns that can help predict future market behavior. The importance of analytical methods lies in their ability to process large volumes of data quickly and accurately, revealing patterns and insights that are crucial for informed financial forecasting and would be difficult for humans to discern.

Building on the knowledge that we have learnt in our lectures, the project is set to dive into the Baijiu(白酒) industry through this project. We aim to depict the landscape of the Baijiu industry through textual analysis, capturing its essence in a concise snapshot. Furthermore, we extends to a thorough dissection of stock prices within this sector, employing a suite of analytical techniques. We try to develop a predictive model which will serve as a cornerstone for understanding and anticipating the stock market's movements within the Baijiu industry.

We choose to analyze Baijiu industry because it’s important for 2 reasons.

-On the one hand, wine holds a unique and profound symbolic significance in Chinese traditional culture.

-On the other hand, Baijiu industry is a key player of China’s stock market. Particularly, Maotai is one of the most valuable public companies, whose performance reflects investor confidence in the broader consumer goods sector.

To analyze wine’s cultural importance, we present a word cloud generated from a list of Tang poems. To do that, we simply split the sentence into words and phrases, and shape them into the image of Libai, the greatest poet. The word cloud shows that wine stands out as the most frequently occurring word. It means that “wine” is extremely important in Chinese culture.

So it’s interesting to investigate wine industry in China’s context. Next we focus on Baijiu industry report in 2023, to see if it’s also important in an economic sense. Basically we use LDA model to extract 7 topics from the report, and visualize the correlation of each word to the topics. As you can see from the heat graph, the baijiu industry excels in quality, market development, with significant growth in international competitiveness.

The cultural and economic significance of Baijiu makes forecasting its industry’s stock prices undeniably important.

The stock price of public firms in Baijiu industry can be obtained from a python library called “tushare”. This library contains all sorts of financial data. Successfully we get daily price information of Baijiu companies from September 20, 2014 to September 20, 2024, all together ten years. For each day, we have open, close, highest and lowest prices, as well as adjustments to make the series more comparable across time.

Here we present raw data for Guizhou Maotai. The stock price of Guizhou Maotai rises significantly from 2015 to 2021, and remains stable at high level even in an age of bear market and economic slowdown.

To get rid of noises in the short run, we also calculate moving average of 50-day, 100-day and 200-day close price, for four prominent Baijiu firms. The results are all similar and robust, which indicates that Baijiu industry provides stability to the stock market.

As the final part of data description, we present daily return of Baijiu firms. As you can see, the return of Guizhou Maotai is very stable.

Since the data high-dimensional characteristics on stock price, we use UMAP technique for dimensionality reduction.

How to interpret this result?

The UMAP result looks quite fancy. The figure maps open, close, highest and lowest daily prices to 2-dimensional plane. Each point in the plane is a daily observation of a Baijiu firm. In the plot, certain regions display clear clustering structure, which might correspond to groups of stocks with similar performance, industry sectors, or some shared characteristics. The large yellow area in the upper right may represent a dense group of stocks with similar characteristics, while the twisted band-like structure on the left suggests that these stocks exhibit relatively unique features.

We also use Minimum spaning trees to cluster daily return. We represent return rate as the difference of log price and calculate Eclidean distance among log returns. Then the MST function will recieve the distance matrix and form minimum trees. For each cluster, we assign a color. The return series consist of 3 clusters, the third one contain years all the way into the futrue, which validates stock prediction.

We also want to use several stocks to predict the price of a certain stock. In that way, we need to ascertain the corelation of different stocks in Baijiu industry.

**7. Predictive Modeling**

Implementation of Support Vector Machine (SVM) or Boosting algorithms for stock price prediction.

Evaluation of model performance and predictive accuracy.

**8. Conclusion**

Finally we summarize findings and their implications for stakeholders in the Baijiu industry. We also highlight future directions for research and analysis in stock prediction.

**9. References**

Slides of the course.