Data 620 Final Project Proposal

Seung-min Song, Krutika Patel

Wine Quality Analysis Using Social Network Analysis and Text Mining

**Abstract:**

This study aims to explore the production of different types of wines in various regions, along with the quality assessment and consumer evaluations of wines in those regions. By integrating social network analysis and text-mining techniques, we aim to extract key information related to wine production, quality, and consumer perceptions from textual data. Network analysis will help us understand the relationships between these factors and how they influence wine pricing. The findings of this research will provide valuable insights for developing future wine production strategies and enhancing our understanding of the dynamics of the wine market.

**Research methods:**

**1. Analysis Data:**

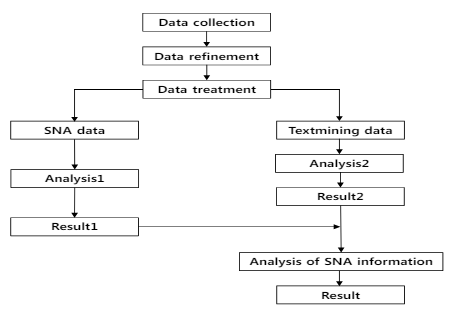
The analysis data utilized in this study are sourced from one primary dataset: Wine Reviews Data. The Wine Reviews Data obtained from Kaggle (https://www.kaggle.com/datasets/samuelmcguire/wine-reviews-data?resource=download), provides information on wine ratings, prices, review, varietal, appellation and additional attributes. The analysis data is constructed by extracting and distinguishing key attributes such as varietal, appellation, rating, price, and review from the datasets. These attributes serve as the basis for subsequent analysis.

**2. Analysis Process:**

The analysis process is outlined in Figure 2.1 and involves the following steps:

1. Data Collection: The initial step involves gathering the data required for analysis from the Wine Reviews dataset.
2. Data Refinement and Processing: Subsequently, the collected data is refined and processed in alignment with the research objectives and the chosen analysis techniques, namely 'network analysis' and 'text mining.' This may include data cleaning, transformation, and integration to create a unified dataset suitable for analysis.
3. Analysis Execution: The refined dataset is subjected to analysis using the two selected techniques: network analysis and text mining. Network analysis explores the relationships and structures within the data, while text mining extracts insights from textual content such as reviews and descriptions.
4. Integration of Results: The results obtained from both network analysis and text mining are synthesized to derive the final findings. Information derived from the network analysis, such as key relationships between wine attributes, is combined with insights obtained through text mining, such as keyword extraction from reviews.

By following this comprehensive analysis process, this study aims to uncover patterns and insights within the wine data that can inform future production strategies and enhance understanding of wine quality and pricing dynamics.



[Figure 2.1: analysis process]