Project Report

Bruno Fernandes Iorio, Beeverly Gourdette

October 8, 2025

1 Problem Framing

In this project, we will study features present in classic literary works and how they intertwine with cultural elements. Specifically, we want to tackle the following questions:

- What are the features in famous literary pieces responsible for their popularity?

- How does culture influence these tastes within certain contexts?

Examples of cultural elements are predominant religions, key historical events (e.g. colonization, wars, etc...), heritage (e.g. immigration), recurring themes (e.g. family, religion...), language, etc.

We can recognize many reasons why these questions are relevant: to understand market necessities of bookstores and publishers, for instance. However, the motivation for the choice of this topic is to study and understand different cultures, learn which cultural aspects shape literary taste and how literary works are perceived across different societies.

2 Strategies

Although we still haven't fully decided which precise resources we will use, we are aware of the existence of many heavily annotated datasets online, such as the CMU Book Summary Dataset, which provides annotated information about books, such as plot, genre, publication date, etc... Another possible source of data would be book reviews online that can be found on Google, which would be used to determine what people from different backgrounds feel about different books. This can be mixed with relevant cultural information which we could likely extract from datasets online such as D-PLACE (for cultural traits), or UNESCO datasets.

We will work a lot with textual feature extraction, which can be a hard task sometimes. To overcome this issue, we consider using pretrained text embedding models, for instance, `Qwen3-Embedding-0.6B`. However, we also may use simpler techniques for data processing, such as "bag-of-words," if it fits our necessities.