**TOPIC MODELLING**

Topic models are active research area in text mining, machine learning, information retrieval, etc. Most of the current statistical topic modeling methods, such as Probabilistic Latent Semantic Analysis (PLSA) and Latent Dirichlet Allocation (LDA). They are used to build models from unstructured text and produce a term-based representation to describe a topic by choosing single words from multinomial word distribution. There are two main weaknesses. First, popular, or common words are different topics, often causing ambiguity for understanding the topics; Second, lack of consistent semantics for single words to be represented correctly. To address these problems, research proposes a model (A Two-Phase Method for Constructing Topic Model, TPMTM) that combines statistical modeling (LDA) with frequent pattern mining and produces better presentations of rich topics and semantics. Empirical evaluation shows that the results of the proposed model are better than LDA.

We first describe and structure these topics, and then further show how the topic focus has evolved over the last two decades. Our study thus provides a structured topography for finance researchers seeking to integrate machine learning research approaches in their exploration of finance phenomena. We also showcase the benefits to finance researchers of the method of probabilistic modeling of topics for deep comprehension of a body of literature, especially when that literature has diverse multidisciplinary actors.

This project mainly focuses on which type of modelling technique can be used for data. We will experiment random datasets across various models and test the accuracy. We can measure the performance of the model by various metrics like root mean squared error, probabilistic error etc.,

**Flow of Project:**

1) Data Collection

2) Data Pre-Processing

3) Exploratory Data Analysis

4) Implement various topic modelling techniques

5) Predict the accuracy