MLT-06- Natural Language Processing and Sentiment Analysis in Trading

Overview

In this lecture, faculty will be covering concepts related to alternative data, NLP concepts and their applications to financial services industry

Key takeaways from the lecture:

- Overview of Alternative Data: Sources, data formats, storage and retrieval choices
- How does one recognize relevant entities in unstructured data?
  - Entity Disambiguation via Permanent Identifier
  - Hands-on exercise
- Understanding RDF and Knowledge Graph
- Tagging Unstructured Data with relevant metadata
  - Hands-on exercise
- Using spaCy for common Text processing tasks
- Understanding Topic Modeling and Topic Classification
  - Evolution of Topic Modeling
  - Hands-on exercise using financial news data
- Understanding Machine Readable News - Programmatic consumption of news
  - Hands-on exercises
- Machine Readable News in the Financial Industry: Sample In-Production use cases
- Sentiment Data in the Financial Industry: Sample In-Production use cases

Instructions to the participants before the session (Click the below link)

Link: file:///C:/Users/Administrator/Downloads/New%20folder/instructions.html

Pre-requisites

- Register at https://developers.refinitiv.com/
- Register at https://permid.org/ and have the API key ready before the session. We will be using the API key for the PermID demo and Intelligent Tagging Demo
- The session would aim to cover most of the content mentioned in the slide deck. If time permits, a few additional aspects related to Sentiment analysis would also be discussed.
- Download sample data from http://bit.ly/3Pr9HG5
- The session would aim to cover most of the content mentioned in the slide deck. If time permits, a few additional aspects related to Sentiment analysis would also be discussed.
- Good understanding of Python as all the hands-on exercises will be based on Python ecosystem
- Prior exposure to Machine Learning principles
- A basic understanding of Equity Markets and Commodity markets as the use cases will demonstrate a few assets from each of the above markets

Practical use of the topics learned in this session

- You will understand the kinds of alternative data that is being used increasingly by various types of hedge funds, HFT players and asset management companies

- You will understand the basic problems that one encounters in typing together various fragments of unstructured data
- You will get to know some of the off-the-shelf resources available to help you disambiguate entities and tag unstructured data
- You will understand the basic building block of graph data model, i.e. RDF
- You will understand the various aspects of Topic Modeling
- You will get a deep dive in to Machine Readable News, that helps one to programmatically consume news
- Based on the combination of concepts from Alternative data, NLP and Machine-Readable News, you will be able to assess the usefulness of various alternative datasets for your trading and risk management use cases

Downloadable Files:

- MLT-06-Files.zip

Recommended time for the session & related coursework: 9 hours