

## CAPSTONE PROJECT SUMMARY

[AUTOMATED TEXT LABELER FOR FINANCIAL NEWS]

AYTAJ A. AHMADOVA  
MS in BUSINESS ANALYTICS

23.06.2020

# Table of Contents

**SUMMARY ..... 3**

**PROBLEM..... 3**

**SOLUTION ..... 3**

**TOOLS..... 3**

**IMPROVEMENT AREAS..... 4**

# Summary

The main objective of the project is to automate the news labeling process, to create interactive dashboards for easier data visualization, and to create informative slides illustrating competitors' movements based on news.

## Problem

Currently, the team members of Private Credit have to collect relevant news for analysis from different data points. They have to read each news to identify whether the news is relevant for market analysis or not. This produces significant delays in the analysis part.

## Solution

The collection of the news from many data points into one database and classification related to the Private Credit sector will help the team to save time, thus increasing the efficiency.

Additionally, automatic text labeling can offer company effective information management that is independent of subjective criteria of categorization, which can lead to inconsistent analysis.

There are two main advantages of the project: the scope size and analysis.

**Scope Size.** The same philosophy can also be found in news platforms. You use keywords to filter down the number of searches. However, the coverage of the news collected by the team is broader. The team collects news from multiple platforms, which gives us a chance to analyze the broader scope rather than focusing on a single source. The broader scope can also allow us to compare different news sources to see which strategies they put a heavier emphasis on.

**Analysis.** Interactive dashboards help us to see the overview of the data in a glance and decide on which competitor movement we should focus on. Once the competitor is chosen based on the provided data, it takes less time to create a competitor movement timeline based on the highlighted news texts.

## Tools

As a main tool, *Python* was used to create the automation processing tool. The labeller was created as JSON file and parsed into the functions as key input. The *NLP* process was applied using several methods such as regular expressions.

The output of the code was an Excel file containing unnested competitor names, investment strategies, timeline, and a document source. To visualize the data, interactive *Tableau* dashboards were created.

## **Improvement Areas**

*The different classification used by different news platforms:* One of the potential problems can be the content of the news used by different news platforms. For instance, some sources classify Private Credit as a subcategory of Private Equity, while others treat them as two different product categories with their own unique strategies.

*Misleading wording in news content:* Another problem is related to the content of the news. To elaborate further, the news can name a specific industry player, which is important for the company to categorize it as a competitor. However, looking at the content one can understand that the name of the competitor is just mentioned as an example and the news itself is not about this particular competitor. Without human interaction, this can harm the analysis used by the team.

One way to address these problems is to build a complex NLP algorithm, which will require lots of news content to train the machine. Considering the scope and technicalities of the project, this process will rather be done manually via extending the dictionary file.