**Data Science Project Report**

**1. Principal Investigator**

Kasi Varma Hasthi: [hasthik1@newpaltz.edu](mailto:hasthik1@newpaltz.edu) / hasthikasivarma@gmail.com

**2. Title of Project**

"Tech Layoffs from 2019 - 2023"

**3. Introduction**

3.1 Project Motivation

The driving force behind the "Tech Layoffs 2019-2023" project is the compelling need to dissect and understand the multifaceted factors that precipitated significant workforce reductions within the global technology sector during this critical period. The project's motivation stems from the overarching goal to unravel the complex tapestry of reasons behind these layoffs, ranging from economic downturns to technological disruptions and strategic transformations within tech companies. By delving deep into these root causes, the project aims to unearth invaluable insights that could illuminate the fundamental drivers steering these pivotal decisions. Utilizing advanced algorithms and comprehensive data analysis techniques, this undertaking aspires to provide a comprehensive understanding of the intricate dynamics that governed the substantial workforce reductions observed within the tech industry from 2019 to 2023.

3.2 Aims and Objectives

The main objective of this project is to complete a comprehensive analysis of the tech layoffs that affected the industry between 2019 and 2023. Identifying the root causes, understanding the primary reasons behind tech layoffs, such as economic downturns, over hiring, technological disruptions, or strategic shifts within companies. This objective helps pinpoint the underlying factors driving workforce reductions. This objective will be achieved by designing a MapReduce algorithm which will be able to handle the large number of datasets we will use. We intend to use the following classification algorithms, decision tree and random forest.

**4. Background/History of the Study**

The "Tech Layoffs 2019-2023" project will unfold against the backdrop of an evolving and challenging era in the global technology sector. From 2019 to 2023, this industry grappled with a convergence of economic and societal factors that reshaped its landscape. Notably, the emergence of the COVID-19 pandemic in early 2020 cast a long shadow over the industry, accentuating the economic pressures already in play. The pandemic accelerated digital transformation, shifting consumer behaviour and business operations worldwide. Tech companies faced the dual challenge of adapting to this rapid digitalization while contending with uncertainties stemming from a global economic slowdown, higher interest rates, and volatile foreign exchange markets.

In this context, the "Tech Layoffs 2019-2023" dataset will be conceived to capture the profound effects of these forces on the tech workforce.

**5. Approach and Implementation**

The "Tech Layoffs 2019-2023" dataset will undergo a well-structured and thorough analysis to uncover insights into reductions in the global technology sector's workforce in the coming critical years. We will commence with an extensive data collection process, gathering information from various sources, including official company reports, reputable news outlets, and government records. This data compilation will encompass detailed layoff information within the broader economic context, with a specific focus on the substantial impact of the COVID-19 pandemic.

To guarantee data accuracy and consistency, we will undertake a rigorous phase of data cleaning and pre-processing. This phase will involve handling missing values, addressing outliers, and standardizing data formats to ensure that the dataset is robust and prepared for in-depth examination.

The exploratory data analysis (EDA) phase will play a pivotal role in gaining an initial understanding of the dataset. We will employ descriptive statistics and data visualization techniques to uncover trends, patterns, and potential outliers. These initial findings will serve as the foundation for subsequent, more focused analyses.

Hypothesis testing and advanced statistical methods will be employed to formulate and investigate specific research questions. These questions will include assessing relationships between economic indicators, evaluating the impact of COVID-19 on layoff trends, and identifying factors that influence the severity and frequency of layoffs.

A dedicated time-series analysis will be carried out to reveal temporal trends and patterns within the dataset. This analysis will aim to capture long-term workforce reduction patterns and cyclic variations, providing insights into the evolving dynamics of tech layoffs over time.

To enhance the analysis, we will utilize sentiment analysis on textual content, such as news articles and press releases associated with workforce reductions. This sentiment analysis will enable us to evaluate the public's outlook, providing insights into how external stakeholders perceive layoffs and how these sentiments could influence stock prices, brand perception, or public policy.

In summary, this comprehensive study aims to provide a holistic understanding of the complex dynamics of tech layoffs in the years 2019 to 2023.

**6. Experiment Results and Discussion**

(Figure 6.1):

A graph of different colored lines

Description automatically generated

Figure 6.1 shows yearly trend for the whole world shows that:

* Layoffs were generally higher in 2020, then they dropped in 2021 and skyrocketed in 2022.
* Initially Transportation and travel industries got affected more as shown by the layoffs in 2020 but they somewhat recovered later.
* Retail, Consumer and Food industries faced sharp increase in layoffs.

(Figure 6.2)

A comparison of a pie chart

Description automatically generated

(Figure 6.3)

**A pie chart with different colored circles

Description automatically generated**

(Figure 6.4)

A graph of different colored lines

Description automatically generated

Figure 6.4 shows yearly layoff trend for united states shows that,

* Layoffs were generally higher in 2020, then they dropped in 2021 and skyrocketed in 2022.
* Initially Transportation and Travel industries got affected more as shown by the layoffs in 2020 but they somewhat recovered later.
* Consumer, Retail, Real Estate, Healthcare, Food and Fitness industries face sharp increase in layoffs in 2022 compared to 2020.
* Yearly layoff trend for the whole world matches with that of yearly trend in USA in the sense that layoffs were higher in 2020, then they dropped in 2021 and skyrocketed in 2022. This may be because majority of data in the dataset is about USA.

(Figure 6.5)

A graph of different colored lines

Description automatically generated

Figure 6.5 shows yearly layoffs trends excluding United States data shows that

* Layoffs were generally higher in 2020, then they dropped in 2021 and greatly increased in 2022
* Initially Travel industry got affected more as shown by the layoffs in 2020 but they recovered very well since then
* Surprisingly Education and Retail industries faced sharp increase in layoffs in 2022 compared to 2020
* Travel and Finance industries performed well in 2022 compared to 2020
* Yearly layoff trend for the whole world (minus USA data) matches somewhat with that of yearly trend in USA in the sense that layoffs were higher in 2020, then they dropped in 2021 and are generally higher in 2022.

(Figure 6.5)

A graph with different colored bars

Description automatically generated

Figure 6.5 shows:

* India suffered most layoffs in Education and Retail Industries.
* India performed well in most industries in 2022 compared to 2021.

(Figure 6.6)

A graph of different colored lines

Description automatically generated

Figure 6.6 Yearly layoff trends excluding United States and India data shoes that:

* Travel industry was badly affected by layoffs in 2020 but it improved greatly in 2022.
* Food, Crypto and Retail industries are badly affected by layoffs in 2022 compared to previous years.
* Surprisingly consumer industry performed well in 2022 compared to 2021.

(Figure 6.7)

A screenshot of a computer program

Description automatically generated

Running through Hadoop MapReduce using layoffs.csv and main.py file.

(Figure 6.8)

A screenshot of a computer

Description automatically generated

After execution, our output file is created in the Hadoop localhost and it is accuracies.json file.

As our dataset belongs to classifier, we decided to use classification models like decision tree classifier, random forest classifier and bagging classifier.

**7. Conclusion**

**A screenshot of a computer

Description automatically generated**

We trained our dataset with three different models,

* Decision tree classifier
* Random forest classifier
* Bagging classifier

From the above figure, we can say that Decision tree and Bagging classifier have slightly highest accuracy with 93% and 94%, than decision tree with 82%.

And here are some of the conclusions we discovered using machine learning model and layoffs dataset:

* United States suffered most layoffs followed by India, compared to the rest of the world. More than 75% of the employees laid off since 2020 belong to USA
* Consumer industry in USA is hit hardest with layoffs in 2022 while Food industry is affected most outside USA in 2022.
* Transportation and Food industries are affected by layoffs globally.
* Travel industry faced considerably less layoffs globally in 2022 compared to 2022.

**8. References**

* <https://layoffs.fyi/> - Original Dataset.
* <https://techcrunch.com/2023/09/19/tech-industry-layoffs-2023/> - Article.
* <https://www.kaggle.com/datasets/swaptr/layoffs-2022/data> - Kaggle Dataset.
* <https://www.youtube.com/watch?v=H999fIuymqc> – Hadoop installation.