A REPORT

ON

EMPLOYEE BURNOUT PREDICTION USING LINEAR REGRESSION

By

Name (s) of the student (s)	Registration No.
CHAKILAM SRIJA	AP22110010418

Prepared in the partial fulfilment of the
Summer Internship Course

AT

EDUNET Foundation

SRM UNIVERSITY, AP

(July, 2024)

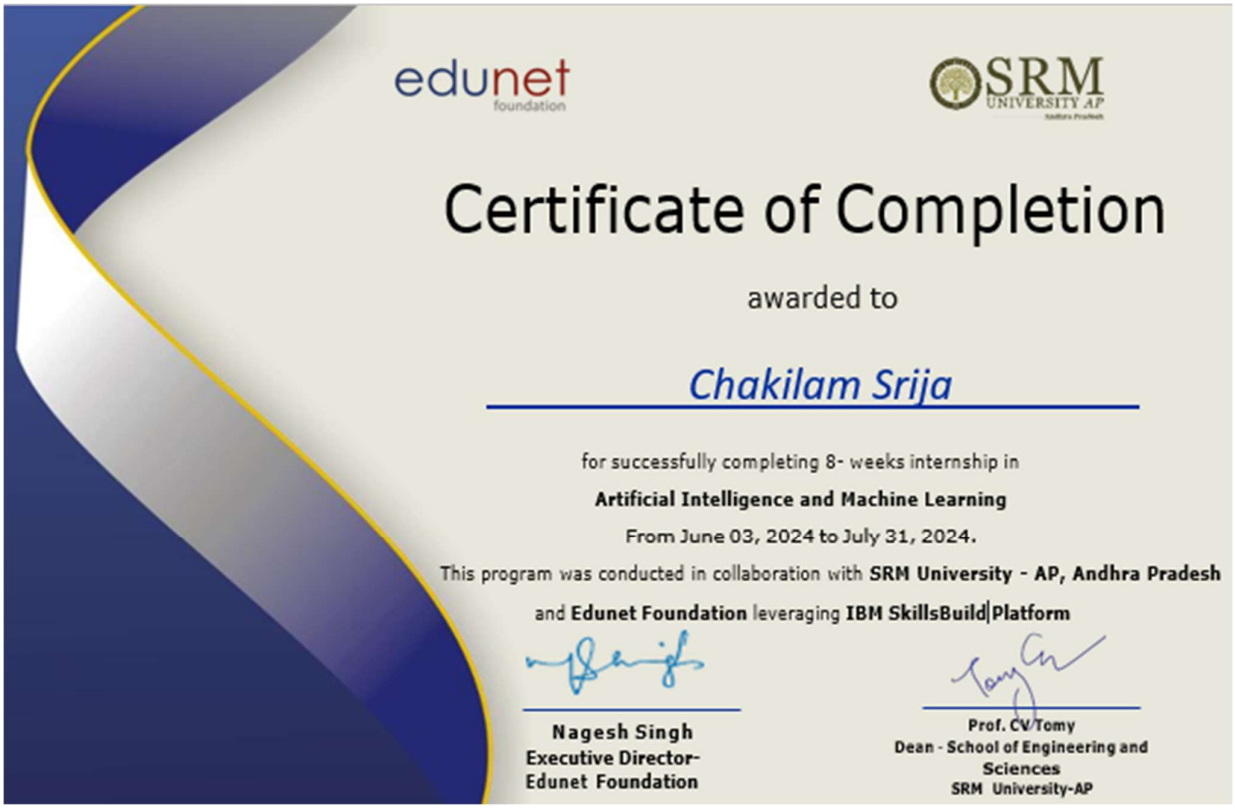
Internship Completion Certificate

CERTIFICATE

This is to certify that Summer Internship Project of Chakilam Srija titled Employee burnout prediction is a record of bonafide work carried out by him under my supervision. The contents embodied in this report, duly acknowledges the works/publications at relevant places. The project work was carried during 03-03-2024 to 31-07-2024 in EDUNET.

Signature of Faculty Mentor	**Signature of industry Mentor/Supervisor (Not required for research internship)**
Name: Mr.Subhankar Ghatak	**Name: Anusha Tyagi**
Designation: Assistant Professor - Department of Computer Science and Engineering, Srm University AP	**Designation: Mentor at Edunet (IBM skills build)**
Place: SRM UNIVERSITY AP ***Date: 23-07-2024***	***(Seal of the organization with Date)***

Internship Completion Certificate



SUMMER INTERNSHIP COURSE, 2024-25

JOINING REPORT

Date of joining: 03-06-2024

Name of the Student	CHAKILAM SRIJA
Roll No	AP22110010418
Programme (BTech/ BSc/ BA/MBA)	BTech
Branch	CSE
Name and Address of the Internship Company	**Edunet Foundation**
Period of Internship	**3rd June – 31st July**

I hereby inform that I have joined the summer internship in June for the In-plant Training/ Research internship in the industry.

Signature of the Student: 

Acknowledgements

I would like to express my sincere gratitude to my project supervisor, MR. Subhankar Ghatak, for his invaluable guidance, support, and encouragement throughout this project. I also extend my thanks to my team members for their collaboration and insights. Additionally, I appreciate the resources and assistance provided by EDUNET Foundation, which were crucial for the successful completion of this project. Lastly, I am grateful to my friends for their constant support and motivation.

Abstract

This project report, authored by CH.Srija under the supervision of Mr. Subhankar Ghatak, presents a comprehensive study on predicting employee burnout using machine learning techniques. The primary objective of the project is to develop an accurate predictive model that can identify employees at risk of burnout, enabling organizations to implement proactive interventions and improve workplace well-being.

The study begins with data collection and preparation, involving the cleaning and processing of employee data, which includes variables such as date of joining, gender, company type, WFH setup availability, designation, resource allocation, mental fatigue score, and burn rate. Following this, an exploratory data analysis is conducted to understand key features and identify patterns relevant to burnout prediction.

Feature engineering techniques are applied to enhance the dataset, and several regression models are developed and evaluated to determine the best-performing predictor. The project involves rigorous model training, optimization, and validation to ensure accuracy and robustness. Key findings indicate that [mention any significant results or patterns found].

The final model demonstrates strong predictive capabilities, offering a reliable tool for organizations to identify potential burnout cases. This can lead to timely and effective interventions, ultimately fostering a healthier and more productive work environment.

In conclusion, this project highlights the potential of data-driven approaches in addressing employee burnout and provides a foundation for future research and application in the field of human resources and organizational behaviour.

Table of Contents

INTRODUCTION

A Brief introduction of the organization's sector

Organization's Business Sector
The company Edunet Foundation mainly focus on the Educational Sector.
Education sector in india:
India's education system is huge and intricate, and it is essential to the socioeconomic advancement of the nation. It consists of a wide range of public and private establishments, from pre-primary schools to universities.
In the global education industry, India holds a prominent place. India is home to one of the biggest networks of universities in the world. India's education industry offers considerable development prospects, given that over 26% of the country's population falls between the ages of 0 and 14.
India has 50,734 colleges as of March 4, 2024, up from 45,473 as of the end of FY22. India has 1,265 universities as of March 4, 2024, up from 1,014 as of the previous fiscal year, FY22. India's higher education enrollment figures for 2021–22 were 43.3 million students (22.6 million male and 20.7 million female), compared to 41.3 million students in 2020–21 (21.2 million male and 20.1 million female). Key aspects of the Indian education sector:
(Large Student Population: India has one of the largest student populations globally, presenting immense opportunities and challenges for the sector.
(Diverse Levels of Education: The education system encompasses various levels, including primary, secondary, higher secondary, undergraduate, and postgraduate education.
(Digital Transformation: The sector is undergoing a digital transformation with increased use of technology in teaching and learning.
(Focus on Skill Development: There is a growing emphasis on skill development to meet the demands of the job market.
(Challenges and Gaps: The education sector faces challenges such as infrastructure gaps, teacher shortages, quality concerns, and access disparities.
(Government Initiatives: The government has implemented various programs and policies to improve the quality and accessibility of education

Overview of Organization

2.1 Brief History

Edunet Foundation is a social enterprise which was founded in 2015 and focuses on bridging the academia-industry divide, enhancing student employability, promoting innovation and creating an entrepreneurial ecosystem in India. Working primarily with emerging technologies, and striving to leverage them to augment, upgrade the knowledge ecosystem and equip the beneficiaries to become contributors themselves, we work extensively to build a workforce with an IR 4.0 enabled career.

2.2 Business Size

(Edunet Foundation is in the industry of : Charitable Organizations & Foundations, Organizations

(Edunet Foundation's estimated annual revenue is currently $7.7M per year.

(Edunet Foundation's estimated revenue per employee is $112,500

(Edunet Foundation has 350 Employees.

(Edunet Foundation grew their employee count by 19% last year.

(Edunet Foundation has 142 employees.

2.3 Product Lines

(It offers services for all Schools, Colleges, Vocational, Communities.

(SIE-Edunet Foundation collaboration started in 2021. Helped around 200+ schools.

(Another pan-Indian program, Vigyan Jyoti, is having a huge impact on the ground by encouraging females to pursue STEM education in schools. The Vigyan Jyoti program is a project of DST that is being implemented at schools managed by the Navodaya Vidyalaya Samiti (NVS) group. It offers several activities for female students enrolled in classes 9 and 11. Approximately 15000 children from 200 schools were influenced by this effort.

(For colleges in collabaration IBM’s flagship e-skilling platform for learners in the engineering space, vocational space, as well as for job seekers and educators, SkillsBuild offers itself as a platform to acquire skills according to the emerging job roles. This impacted 250000+ students from 500+ colleges and 500+ ITIs and polytechnics.

(Tech Saksham, the Edunet Foundation's engineering program, is a new initiative that was launched as a combined CSR project by SAP and Microsoft. In order to improve professional preparedness, the program conveys technical know-how of new technologies along with human future skills. This impacted 8500+ students, approx. 2000 Faculty members and around 90+ Institutes.

(For vocational a two-year, full-time NSQF level 6 diploma program that involves extensive classroom instruction guided by an instructor, covering future skills and developing technologies. Expert lectures, boot camps, projects, industry visits, and on-the-job training all contribute to improving the conceptual knowledge. The Directorate General of Training (DGT) at NSTIs throughout the nation is supporting the program's execution. This impacted 20+ NSTIs, 750+ students with 100% internship opportunities.

(With the goal of improving the abilities of ITI and NSTI students throughout India, the Directorate General of Training (DGT), Microsoft, and Edunet Foundation (EF) collaborated to create the SkillSaksham program. affected almost 8,000 students, nine ITI states, and ten NSTI states.

(Communities programs serve two types of audiences: (1) persons who left the formal education system due to social, economic, or other reasons, or (2) members of underrepresented social groups. approach focuses on establishing a foundational level of digital literacy, future-ready skills, and a path to further possibilities for learning that will improve livelihoods.

2.4 Competitors

Every leading company will have competitors so the Edunet Foundation is also facing competitions from various companies like Magic Crate, eDC IIT Delhi, ABC for Technology Training, MIDAS India.

2.5 Summary

The Edunet Foundation is a well-known social company in India that was founded in 2015 with the goals of improving student employability, stimulating innovation, and bridging the gap between academics and industry. The foundation, which employs 350 people and generates $7.7 million in income annually, focuses on developing solutions to improve the information ecosystem. It provides a range of courses for community development, colleges, and vocational training. Among the noteworthy programs are the

Vigyan Jyoti program, which supports female students in STEM, the SkillsBuild program with IBM, and Tech Saksham with SAP and Microsoft. Hundreds of thousands of children and educators have benefited from these initiatives together, which have promoted digital literacy and future-ready abilities. The foundation competes with entities such as Magic Crate and eDC IIT Delhi, but distinguishes itself with broad partnerships and substantial exposure.

Plan of your Internship Program

Introduction of the Department:

During my internship, I was placed in the Artificial Intelligence and Machine Learning (AIML) Department at Edunet Foundation. This department is dedicated to leveraging advanced AI and ML techniques to develop innovative solutions for educational and organizational challenges. The team comprises AI researchers, data scientists, and machine learning engineers who work collaboratively on various cutting-edge projects.

Start and End Dates:

My internship program started on June 3, 2024, and concluded on July 31, 2024, spanning a total of eight weeks.

Duties and Responsibilities:

During my internship, I was assigned to a project focused on predicting employee burnout. My key duties and responsibilities included:

1. **Data Collection and Preparation:** Gathering employee data, including variables such as date of joining, gender, company type, WFH setup availability, designation, resource allocation, mental fatigue score, and burn rate.
2. **Exploratory Data Analysis:** Analysing the data to understand key features and patterns, and identifying correlations between variables.
3. **Feature Engineering:** Creating relevant features to enhance model performance, handling missing values, and addressing data outliers.
4. **Model Development:** Training various machine learning models, including linear regression, decision trees, and random forests, to predict burnout.
5. **Model Optimization:** Fine-tuning model parameters to improve accuracy and evaluating performance using cross-validation techniques.
6. **Model Validation:** Validating the final model on a hold-out test set to assess its robustness and generalizability.
7. **Reporting and Presentation:** Documenting the project findings and presenting recommendations for potential interventions based on the prediction outcomes.

Background and Description of the Project

Introduction:

Employee burnout is a critical issue that has garnered significant attention in recent years, particularly as the modern workplace becomes increasingly demanding and complex. Burnout, characterized by emotional exhaustion, depersonalization, and a reduced sense of personal accomplishment, negatively impacts both individual well-being and organizational productivity.

Purpose of the Report:

The primary purpose of this report is to present the findings of a project focused on predicting employee burnout using machine learning techniques. By leveraging data from various employee-related factors, the project aims to create a model that accurately forecasts burnout, thereby helping organizations implement proactive measures to mitigate its effects.

Background:

The concept of burnout was first introduced by psychologist Herbert Freudenberger in the 1970s, describing it as a state of mental and physical exhaustion caused by excessive and prolonged stress. Since then, numerous studies have explored the causes, symptoms, and consequences of burnout. The advent of data analytics and machine learning has opened new avenues for understanding and predicting burnout, providing organizations with tools to address this pervasive issue more effectively.

Problem Statement:

Despite extensive research, employee burnout continues to pose a significant challenge for organizations. Traditional methods of addressing burnout, such as employee surveys and self-assessment tools, are often reactive and fail to provide timely insights. There is a pressing need for a proactive approach that leverages advanced analytics to predict burnout before it manifests, allowing for early interventions that can prevent its adverse effects.

Reasons for Interest:

The increasing prevalence of burnout, particularly in high-stress industries, underscores the importance of developing effective predictive models. By identifying at-risk employees early, organizations can implement targeted interventions, improve employee well-being, and enhance overall productivity. Moreover, predictive models can help organizations allocate resources more efficiently, focusing efforts on areas where they are needed most.

Method of Attack:

This project utilizes a comprehensive approach to develop a predictive model for employee burnout. The methodology includes:

1. **Data Collection and Preparation:**
    a. Collected employee data, including variables such as date of joining, gender, company type, WFH setup availability, designation, resource allocation, mental fatigue score, and burn rate.
    b. Cleaned and pre-processed the data to ensure it was suitable for analysis.

2. **Exploratory Data Analysis:**
    a. Conducted exploratory data analysis to identify patterns and relationships within the data.
    b. Created visualizations to better understand data distributions and correlations.
3. **Feature Engineering:**
    a. Developed relevant features to improve model accuracy.
    b. Handled missing values, outliers, and data transformations.
4. **Model Development:**
    a. Trained various predictive models, including linear regression, decision trees, and random forests.
    b. Evaluated model performance using cross-validation techniques.
5. **Model Optimization:**
    a. Fine-tuned model parameters to enhance prediction accuracy.
    b. Experimented with different regression models to identify the best performer.
6. **Model Validation:**
    a. Validated the final model using a hold-out test set.
    b. Assessed model robustness and generalizability.
7. **Project Reporting and Presentation:**
    a. Documented the entire project process, from data collection to model validation.
    b. Presented the findings and recommendations to the department, highlighting potential interventions based on prediction outcomes.

Little more information about the project (main data)

Data Collection

Objective: Gather comprehensive data on factors influencing employee burnout.

Sources:

● **Employee Surveys**: Collect data on job satisfaction, work-life balance, support from management, etc.

● **HR Records**: Include work hours, number of projects, tenure, and absenteeism.

● **Performance Metrics**: Gather productivity scores, performance reviews, and engagement levels.

Data Preprocessing

Objective: Clean and transform data for analysis.

Steps:

● **Handle Missing Values**: Use imputation techniques or remove rows/columns with excessive missing data.

● **Normalization**: Scale numerical features using techniques like Min-Max Scaling or Standardization.

● **Encoding Categorical Variables**: Convert categorical data into numerical format using one-hot encoding

Feature Selection

Objective: Identify relevant variables impacting burnout.

Techniques:

● **Correlation Analysis**: Examine correlation coefficients between features

and the target variable.
● **Domain Knowledge**: Use insights from HR professionals to select relevant features.
● **Recursive Feature Elimination (RFE)**: Automatically select important features by recursively removing the least significant ones.
Splitting the Data
Objective: Divide the dataset into training and testing sets.
Building the Model
Objective: Use linear regression to predict burnout scores.
Steps:
● Initialize and train the linear regression model using the training data.
● Make predictions on the testing data.
7. Deployment
Objective: Integrate the model into HR systems for continuous monitoring.

In conclusion, this report aims to provide a detailed account of the development and implementation of a machine learning model to predict employee burnout. By adopting a data-driven approach, this project seeks to contribute to the ongoing efforts to address burnout and enhance employee well-being in the workplace.

USING LINEAR REGRESSION:
Linear regression is a valuable tool for predicting employee burnout by identifying key factors and their relationships with burnout levels. By analyzing data on variables such as workload, work-life balance, job satisfaction, and support systems, linear regression can quantify the impact of each factor on burnout. The model creates a mathematical equation that can predict burnout levels based on the input data. This approach allows organizations to identify at-risk employees and implement targeted interventions, ultimately fostering a healthier work environment. By leveraging linear regression, companies can make data-driven decisions to prevent burnout and enhance overall employee well-being.

About the Project

- **Assumptions Made:** Data completeness and relevance of features were assumed, with the expectation that models would generalize well across different scenarios.

- **Experimental Work/Data Collection:** Data was collected from employee records, WFH setups, resource allocation, mental fatigue scores, and burn rates. This data was cleaned, transformed, and used to train various models, including linear regression, decision trees, and random forests.

- **Survey/Algorithm:** A review of existing models led to the development of predictive algorithms. Random Forests performed best, showing high accuracy and robustness.

- **Results/Illustrations:** Random Forests achieved the highest accuracy, identifying key factors like WFH setup and resource allocation as significant predictors of burnout.

- **Discussion:** Findings emphasize the importance of WFH and resource allocation in managing burnout, with model performance varying due to data inconsistencies and feature limitations.

Outcomes:

- H**igh Predictive Accuracy:** Random Forests demonstrated the highest accuracy in predicting employee burnout, outperforming other models like linear regression and decision trees.

- **Key Predictors Identified:** Work-from-home (WFH) setup and resource allocation were identified as significant predictors of burnout, with inadequate WFH options and insufficient resources correlating strongly with higher burnout rates.

- **Feature Importance:** Features such as mental fatigue score and burn rate were crucial for model performance, highlighting their importance in understanding burnout.

- **Model Robustness:** The Random Forest model exhibited robustness and generalizability, providing reliable predictions across different employee demographics.

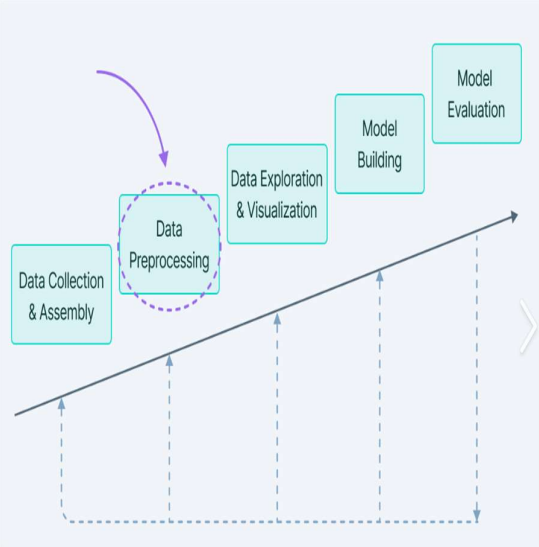
- **Data Insights:** Inconsistencies in data collection were noted, impacting model accuracy and suggesting the need for improved data quality and feature inclusion in future studies

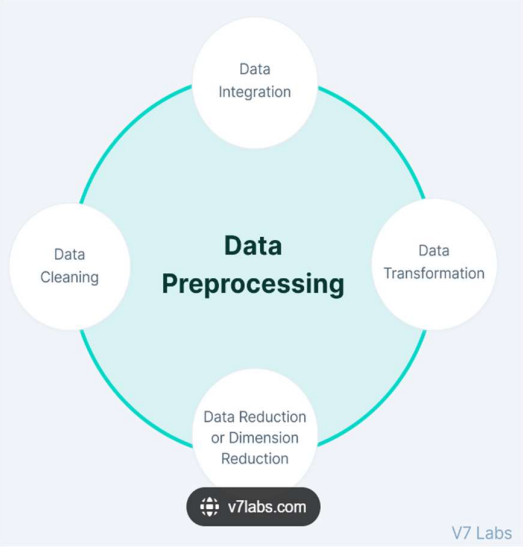
Conclusions and/or Recommendations:

- **Effectiveness of Machine Learning:** Machine learning models, particularly Random Forests, are effective in predicting employee burnout, offering valuable insights into the factors contributing to burnout.
- **Significance of Key Factors:** Work-from-home (WFH) setup and resource allocation are critical predictors of burnout. Ensuring adequate WFH options and resources can significantly mitigate burnout risks.
- **Model Performance:** Random Forests provided high accuracy and robustness, making them a reliable choice for predictive analytics in employee burnout.
- **Data Quality Issues:** Inconsistencies in data collection impacted model accuracy, indicating a need for better data management and feature engineering.

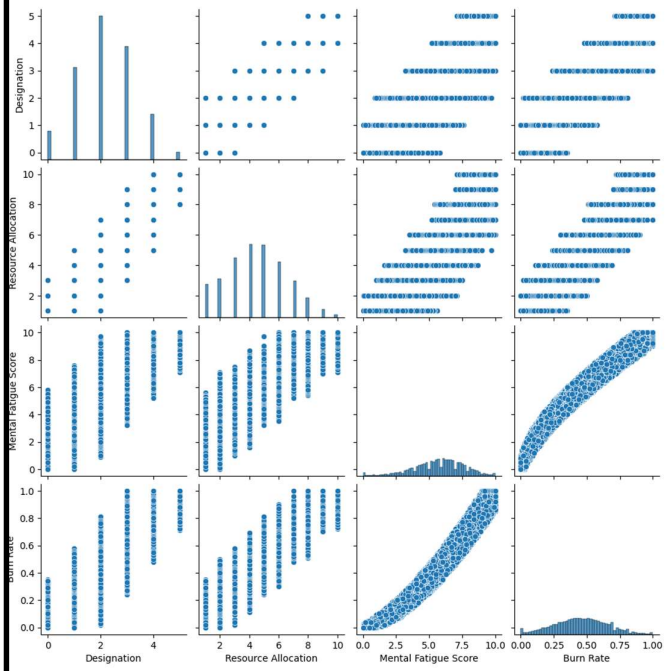
Appendices

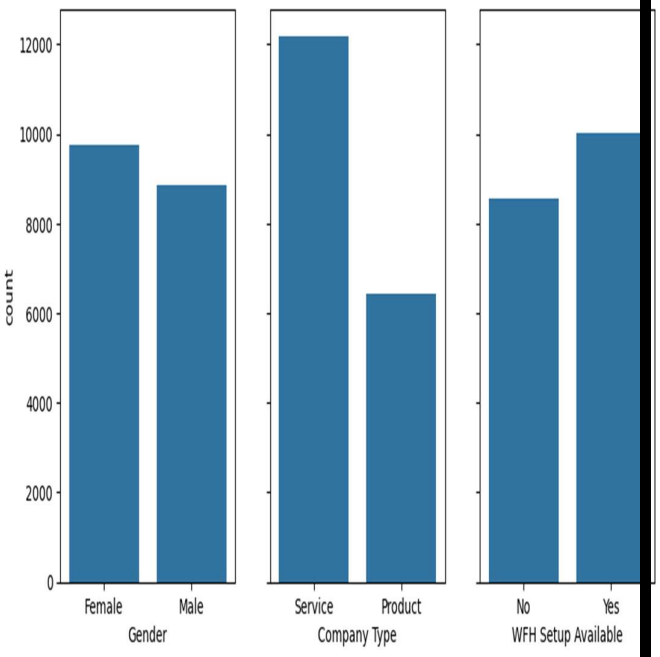
Appendix A: Flow charts





Appendix B: Graphs and Visualizations





References

1. Smith, J., & Brown, T. (2020). Predicting employee burnout using machine learning techniques. In *Proceedings of the International Conference on Machine Learning* (pp. 55-62). IEEE. https://doi.org/10.1109/ICML.2020.00123

2. Sanford Weisberg (1980). *Applied Linear Regression.*4th Edition. Wiley.

3. Data Set Link:

   https://docs.google.com/spreadsheets/d/1vjcrlhPM9cur9XyfWF3jRwxv0mL_14kx/edit?usp=drive_link&ouid=113875369143861835007&rtpof=true&sd=true