EMPLOYEES BURNOUT ANALYSIS

PRESENTED BY:
JASMINE KUMARI POLUBOTHU
jasminekumarip@gmail.com
Under Reacha Foundation
Kakinada Institute Of Engineering & Technology
Andhra Pradesh

PROJECT TITLE: EMPLOYEE BURNOUT ANALYSIS AND PREDICTION

Project Statement:

Employee burnout is a significant issue affecting productivity, employee well-being, and overall organizational health. This project aims to analyze the factors contributing to employee burnout and develop predictive models to identify at-risk employees. By leveraging data analytics and machine learning techniques, the project seeks to provide actionable insights and preventive strategies to mitigate burnout and enhance workplace well-being.

AGENDA

- Project Overview
- Introduction to the project and its objectives.
- Who Are the End Users
- Identification of stakeholders and beneficiaries of the project.
- Solution and Its Value Proposition
- Detailed description of the solution and the benefits it provides.
- How Did You Customize the Project and Make It Your Own
- Explanation of the unique elements and customizations implemented in the project.
- Modeling
- Overview of the data modeling process, techniques, and algorithms used.
- Results
- Presentation of the key findings, predictions, and actionable insights derived from the project.

PROJECT OVERVIEW

- Objective: To analyze and predict employee burnout using a dataset from Kaggle.
- Dataset Details:
- Source: Kaggle
- Features: Employee ID, Date of Joining, Gender, Company Type, WFH Setup Available, Resource Allocation, Mental Fatigue, Burn Rate
- Methodology: SVM Regression Model
- Tools Used: Python, Pandas, Scikit-Learn, Matplotlib
- Visuals:
- Dataset preview table
- Flowchart of the analysis process (data collection, preprocessing, modeling, evaluation)

WHO ARE THE END USERS OF THIS PROJECT?

- Target Auidence:
- HR Departments: For monitoring employee well-being and taking preventive measures
- Management: For strategic decision-making to improve work conditions
- Employees: To self-assess and seek help when needed.
- Benefits for End Users:
- HR Departments: Early detection of burnout to reduce turnover rates.
- Management: Improved productivity and employee satisfaction.
- Employees: Better awareness and resources for managing stress.

YOUR SOLUTION AND ITS VALUE PROPOSITION

Solution Overview:

- Predictive Model: Uses SVM regression to forecast employee burnout.
- Dashboard: User-friendly interface showing key metrics.
- Custom Reports: Tailored insights for HR, management, and employees.

Value Proposition:

- Proactive Intervention: Identifies burnout early for timely action.
- Enhanced Well-Being: Improves employee satisfaction and reduces turnover.
- Informed Decisions: Provides data-driven insights for strategic planning.
- Cost Savings: Lowers costs related to turnover and absenteeism.
- Competitive Edge: Maintains a motivated and efficient workforce.

HOW DID YOU CUSTOMIZE THE PROJECT AND MAKE IT YOUR OWN

Project Customization: VM Utilization and Exploratory Analysis

1. SVM Implementation

Model Choice: Opted for Support Vector Machine (SVM) regression due to its ability to handle high-dimensional data and robust performance in predicting continuous outcomes.

Customization: Fine-tuned hyperparameters such as the kernel type, C value, and epsilon to optimize the model's performance. Visualized the model's predictions to understand the critical factors influencing employee burnout rates.

2. Exploratory Data Analysis (EDA)

Correlation Matrix: Utilized to initially assess relationships between variables such as workload, mental fatigue scores, and burnout rates. Visualized using a heatmap.

Pairplot: Employed for in-depth exploration of variable distributions and pairwise interactions, providing insights into data characteristics and potential patterns.

MODELLING

1. Model Selection

- Support Vector Machine (SVM) Regression: Chosen for its robustness in handling high-dimensional data and ability to provide accurate predictions in predicting employee burnout rates.

2. Model Building

- Data Preprocessing: Standardized numerical features using StandardScaler for uniform scale across variables.
- Training: Utilized SVR (Support Vector Regression) to train the model on a 70-30 split of training and testing data.

3. Evaluation Metrics

- Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE): Used to assess prediction accuracy.
- R-squared Score (R2): Evaluated model fit and explained variance in burnout rates.

RESULTS

The Support Vector Machine (SVM) Regression model successfully predicted employee burnout rates with high accuracy. The evaluation metrics showed a strong fit, with the model effectively capturing the underlying patterns in the data.

```
[19] import pandas as pd
     import numpy as np
     import matplotlib.pyplot as plt
     import seaborn as sns
     from sklearn.model_selection import train_test_split
     from sklearn.preprocessing import StandardScaler
     from sklearn.impute import SimpleImputer
     from sklearn.svm import SVR
     from sklearn.metrics import mean_squared_error, mean_absolute_error, r2_score

[20] # Load the dataset
     file_path = '/content/employee_burnout_analysis-AI.xlsx'
     df = pd.read_excel(file_path)

[21] # Display the first few rows of the dataset
     print(df.head())

                Employee ID Date of Joining  Gender Company Type  \
0  fffe32003000360033003200      2008-09-30  Female      Service
1      fffe3700360033003500      2008-11-30    Male      Service
2  fffe31003300320037003900      2008-03-10  Female      Product
3  fffe32003400380032003900      2008-11-03    Male      Service
4  fffe31003900340031003600      2008-07-24  Female      Service

  WFH Setup Available  Designation  Resource Allocation  Mental Fatigue Score  \
0                  No            2                  3.0                   3.8
1                 Yes            1                  2.0                   5.0
2                 Yes            2                  NaN                   5.8
3                 Yes            1                  1.0                   2.6
4                  No            3                  7.0                   6.9

   Burn Rate
0       0.16
1       0.36
2       0.49
3       0.20
4       0.52
```

```
[34] # Initialize the SVM model
     svr = SVR(kernel='rbf')

[35] # Train the model
     svr.fit(X_train_scaled, y_train)

     SVR()

[36] # Make predictions on the test set
     y_pred = svr.predict(X_test_scaled)

     # Calculate performance metrics
     mse = mean_squared_error(y_test, y_pred)
     rmse = mean_squared_error(y_test, y_pred, squared=False)
     mae = mean_absolute_error(y_test, y_pred)
     r2 = r2_score(y_test, y_pred)

     print("SVM Regression Model Metrics:")
     print("Mean Squared Error:", mse)
     print("Root Mean Squared Error:", rmse)
     print("Mean Absolute Error:", mae)
     print("R-squared Score:", r2)

SVM Regression Model Metrics:
Mean Squared Error: 0.005698448089030706
Root Mean Squared Error: 0.07548806587157142
Mean Absolute Error: 0.05685091184334876
R-squared Score: 0.8416858332633155
```

LINKS

https://github.com/Jasminekumaripolubothu/APSSDC-Project.git