Mental Health Predictor - Capstone Project

Overview

This project is part of the final exam for INSY 8413 - Introduction to Big Data Analytics. It focuses on analyzing and predicting mental health treatment needs among tech workers using Python and visualizing insights using Power BI.

Problem Statement

Can we predict whether a person in the tech industry is likely to seek mental health treatment based on their background, workplace environment, and personal history?

Sector

Health

Dataset

- Source: Kaggle - Mental Health in Tech Survey
- Format: CSV
- Structure: Structured Data
- Preprocessing:
  - Cleaned age values
  - Encoded categorical features
  - Handled null/missing values

Tools Used

- Python (Jupyter Notebook): Data cleaning, exploratory data analysis (EDA), and model building using Random Forest
- Power BI: Created an interactive dashboard for visual storytelling and insights

Key Features Used

- Age
- Gender
- Family History
- Remote Work
- Mental/Physical Consequences
- Supervisor & Coworker Support
- Employer Benefits & Wellness Programs

Machine Learning Model

- Model: Random Forest Classifier
- Evaluation Metrics: Accuracy, Confusion Matrix, Feature Importance

Power BI Dashboard Highlights

- Bar Chart: Mental health treatment by gender
- Pie Chart: Family history distribution
- Line Graph: Age vs. treatment likelihood
- Slicers: Age, support availability, wellness programs
- KPI Card: % of individuals needing treatment

Key Insights

- Family history and lack of employer support are strong predictors of mental health treatment needs
- Individuals aged 20-35 are more likely to seek help
- Remote work flexibility and wellness programs significantly impact mental health outcomes

Files in This Repository

- mental_health_analysis.py - Full Python code for analysis and modeling

- cleaned_mental_health.csv - Cleaned dataset for Power BI dashboard

- README.md - Project overview and instructions

- presentation.pptx - Summary presentation (optional)

How to Use

1. Download the dataset from Kaggle

2. Run the Python script to preprocess and analyze the data

3. Import the cleaned CSV file into Power BI

4. Create or explore the dashboard

5. Present the findings using the PowerPoint file

Author

- Name: Mugisha Igor

- Course: INSY 8413 - Introduction to Big Data Analytics

- Instructor: Mr. Eric Maniraguha

"Work hard with all your heart, as working for the Lord." - Colossians 3:23