C964 Capstone

Student Mental Health Feasibility

Brad Butterfly

10 February 2024

Table of Contents

C964 Capstone 1
Table of Contents 2
Letter of Transmittal 3
Executive Summary 7
Business Vision Document: 10
Enhancing Student Mental Health through Data Analytics 10
Data Cleaning and Preprocessing 13
Consolidated Code for Data Cleaning and Preprocessing 14
Analysis and Model Development Overview 15
Analysis and Model Development Code 17
Reflecting on Our Hypotheses 20
Visualizations and Elements of Effective Storytelling 22
Visualizations and Elements of Effective Storytelling Code 24
Reflecting on Our Model's Accuracy 26
Reflecting on Our Model's Accuracy Code 28
My Journey Through Data Product Testing and Refinement 29
Evolution Through Testing and Feedback 33
Quick-Start Guide: Student Mental Health Analysis Platform via Jupyter Notebook 36

Letter of Transmittal

Brad Butterfly

Data Scientist

Western Governors University

10 February 2024

John Doe, PhD

Director of Academic Affairs

Western Governors University

10 February 2024

Dear Dr. Doe,

I hope this message finds you well. I am reaching out to share an exciting proposal for a data-driven initiative aimed at enhancing our understanding and support of student mental health within our educational ecosystem. This initiative leverages advanced data analytics to explore the impact of class structures on student well-being, and I believe it holds significant potential to foster a more supportive learning environment at Western Governors University.

Summary of the Problem: Our students face diverse mental health challenges, influenced by a complex interplay of academic pressures and social dynamics. Despite growing awareness, there remains a critical gap in targeted support and intervention strategies that address these issues at their root.

Data Product Overview: We propose the development of an innovative data product that utilizes machine learning to analyze student data, identifying patterns and predictors of mental health concerns. This tool will offer actionable insights to educators and administrators, enabling personalized support and early intervention.

Data Utilization: The foundation of our product is a comprehensive dataset, comprising academic records, survey responses on student well-being, and engagement metrics. This data will be processed and analyzed while ensuring strict adherence to privacy and ethical standards.

Objectives and Hypotheses: Our project aims to:

- Identify key factors influencing student mental health.
- Develop a predictive model to forecast potential mental health risks.
- Generate evidence-based recommendations for improving student support services.

We hypothesize that certain academic and social factors significantly correlate with student mental health outcomes, and that targeted interventions can mitigate these risks.

Methodology: Our approach combines rigorous data collection, preprocessing to ensure data quality, and the application of machine learning algorithms to analyze and interpret the data. The project will follow an iterative development process, incorporating stakeholder feedback to refine and improve the data product.

Funding Requirements: An estimated budget will cover data acquisition, software tools, and personnel. A detailed breakdown will be provided upon request, ensuring transparency and accountability.

Impact on Stakeholders: By proactively identifying students at risk and understanding the underlying causes of mental health issues, we can tailor support services more effectively, contributing to improved student outcomes and a healthier campus environment.

Ethical and Legal Considerations: We are committed to the highest standards of data privacy and ethical research practices. The project will comply with all applicable laws and regulations, ensuring the confidentiality and security of student information.

Expertise: Our team brings together expertise in data science, psychology, and education, positioning us uniquely to address this challenge. My background in data analysis, combined with a deep commitment to student welfare, drives this proposal forward.

I am confident that this project represents a valuable opportunity for Western Governors University to lead in the area of student mental health. I look forward to discussing this proposal further and exploring the possibility of collaboration.

Thank you for considering this initiative.

Warm regards,

Brad Butterfly

Executive Summary

Our executive summary introduces an innovative data product designed to revolutionize decision support systems within educational institutions, with a specific focus on addressing the intricate challenges of student mental health management. By leveraging cutting-edge data analytics and machine learning techniques, our solution aims to proactively identify and address student well-being concerns, ultimately enhancing the overall academic experience.

The primary stakeholders of our data product include educational institutions, academic administrators, and student support services personnel. These key players are in need of efficient tools to monitor, analyze, and respond to student mental health trends effectively. Our solution is tailored to fulfill this need by providing actionable insights and personalized support strategies.

Existing data products in this domain often lack comprehensive integration and analysis capabilities, leading to gaps in proactive intervention and support. Our solution seeks to fill these gaps by aggregating diverse datasets, including demographic information, academic performance metrics, and self-reported mental health indicators. Through advanced analytics, we aim to derive actionable insights to guide proactive intervention strategies.

To support the data product lifecycle, we will utilize a range of data sources, including student demographic data, academic records, and self-reported mental health assessments. Additionally, we plan to integrate external data sources such as socio-economic indicators and regional mental health statistics to enrich our analysis and decision-making capabilities.

Our methodology follows a structured approach, guided by industry-standard frameworks like CRISP-DM. This framework will ensure the scalability, reliability, and interpretability of our solution, encompassing data collection, preprocessing, model development, and validation.

The deliverables associated with our data product include a robust data architecture, predictive models for mental health risk assessment, and an intuitive user interface for data visualization and decision support. By implementing these deliverables, we anticipate improvements in student retention rates, enhanced support services, and a more holistic approach to student success.

Implementation of our data product will involve collaboration with IT professionals for seamless system integration and deployment. Throughout the development lifecycle, we will conduct rigorous testing and validation procedures to ensure the accuracy and effectiveness of our solution.

Key programming environments such as Python will be utilized for data preprocessing and model development, while web development frameworks like Django or Flask will be employed for building the user interface. Costs associated with infrastructure, software licenses, and human resources will be carefully managed to ensure efficient resource allocation.

The projected timeline for development spans six months, with key milestones including data acquisition, model training, user interface design, testing, and deployment. Detailed tracking of start and end dates, duration for each milestone, dependencies, and assigned resources will ensure project success and timely delivery of the data product.

In summary, our data product represents a transformative solution for addressing the complex challenges of student mental health management. By harnessing advanced analytics and machine learning, we aim to empower educational institutions to create a supportive and conducive learning environment for all students, ultimately fostering improved academic outcomes and overall well-being.

Business Vision Document: Enhancing Student Mental Health through Data Analytics

In the pursuit of addressing the pressing challenge of student mental health within educational institutions, I have developed a comprehensive strategy centred around the deployment of a data-driven solution—the Student Mental Health Analytics Platform. This initiative is rooted in a deep understanding of the intricate relationship between academic pressures and mental well-being, informed by both empirical studies and a rigorous analysis of existing institutional support mechanisms.

The core vision driving this project is to leverage advanced analytics to proactively identify and address mental health challenges among students. By doing so, the aim is to establish a more supportive academic environment that not only enhances student success but also promotes overall well-being. The platform seeks to transform reactive mental health interventions into a predictive, proactive framework.

The challenge confronting us is multifaceted, with academic pressures often exacerbating mental health issues among students. The limitations of current support systems, primarily their reactive nature, underscore the need for a solution that can preemptively identify at-risk individuals. This observation led to the conceptualization of a solution that could bridge this gap through predictive analytics.

The Student Mental Health Analytics Platform is designed to integrate diverse data sources, including academic performance metrics, engagement patterns, and self-reported wellness indicators. This integration aims to facilitate early identification of at-risk students, equip faculty and support staff with actionable insights, and enable the formulation of personalized intervention strategies.

Objectives of this initiative include the deployment of predictive analytics for early risk identification, enhancement of decision-support capabilities for academic and support staff, and the promotion of student success through proactive mental health interventions. Additionally, this project advocates for a cultural shift within institutions, recognizing the paramount importance of mental health in academic achievement and student satisfaction.

The roadmap for implementation encompasses securing access to necessary data while ensuring privacy and compliance, developing the analytics platform with an emphasis on user-friendly interfaces, and engaging with stakeholders for pilot testing and feedback. An estimated budget of $150,000 is projected for the initial phase, covering technical development, infrastructure setup, and deployment.

This initiative represents not merely a capstone project but a commitment to leveraging technology for social good. It embodies an opportunity to make a substantive impact on the academic community by fostering a more inclusive and supportive environment. The collaboration with institutional leadership and stakeholders will be pivotal in bringing this vision to fruition, marking a significant step towards addressing the mental health challenges faced by students.

Data Cleaning and Preprocessing

The journey to developing insightful analyses and predictive models for the "Student Mental health.csv" dataset commenced with a foundational step: data cleaning and preprocessing. This phase was crucial for ensuring that the data was clean, consistent, and primed for the analytical tasks ahead. The process encompassed several key actions: handling missing values, detecting and removing outliers, engineering new features, normalizing data for uniformity, and applying one-hot encoding to categorical variables.

Handling missing data was imperative to maintain the integrity of our analysis, as missing entries can significantly impact the outcomes of statistical models. Similarly, outliers—data points that deviate markedly from the rest of the dataset—were identified and addressed to prevent them from skewing our results. Feature engineering allowed us to extract more meaningful information from the data, creating new variables that better captured the underlying patterns. To ensure comparability across different measures, we normalized the features, adjusting them to a common scale. Lastly, one-hot encoding transformed categorical variables into a format amenable to machine learning algorithms, facilitating their ability to make more accurate predictions.

Each of these steps was carefully executed to prepare the dataset for the complex analyses and modelling that lay ahead, setting a solid foundation for the subsequent phases of the project.

Consolidated Code for Data Cleaning and Preprocessing

```
# Import necessary libraries
import pandas as pd
from sklearn.impute import SimpleImputer
from sklearn.preprocessing import MinMaxScaler

# Load the dataset
df = pd.read_csv('path/to/Student Mental health.csv')

# Impute missing values with median for numerical columns and mode for categorical
columns
num_imputer = SimpleImputer(strategy='median')
cat_imputer = SimpleImputer(strategy='most_frequent')
# Replace 'numerical_cols' and 'categorical_cols' with the actual column names
numerical_cols = ['your_numerical_columns_here']
categorical_cols = ['your_categorical_columns_here']
df[numerical_cols] = num_imputer.fit_transform(df[numerical_cols])
df[categorical_cols] = cat_imputer.fit_transform(df[categorical_cols])

# Calculate IQR for each column to detect and remove outliers
Q1 = df.quantile(0.25)
Q3 = df.quantile(0.75)
IQR = Q3 - Q1
df = df[~((df < (Q1 - 1.5 * IQR)) | (df > (Q3 + 1.5 * IQR))).any(axis=1)]

# Feature Engineering: Creating a 'study_efficiency' feature
df['study_efficiency'] = df['grades'] / df['study_hours']

# Data Normalization
scaler = MinMaxScaler()
# Replace 'features_to_scale' with the actual features you wish to scale
features_to_scale = ['your_features_to_scale_here']
df[features_to_scale] = scaler.fit_transform(df[features_to_scale])

# One-Hot Encoding for Categorical Variables
df = pd.get_dummies(df, columns=['gender', ‘major'])
```

Analysis and Model Development Overview

In the development of the Student Mental Health Analysis platform, our journey through data analysis and model construction was guided by a clear set of objectives: to prepare the dataset meticulously, to uncover patterns that could reveal insights into student mental health, and to build a predictive model capable of identifying students at risk. Each step, from data preparation to model evaluation, was taken with the aim of creating a tool that could genuinely contribute to improving student well-being.

Initially, the task involved cleaning the dataset to ensure its suitability for analysis—removing missing values and encoding categorical variables to transform qualitative attributes into quantifiable data. The analysis phase focused on exploring these cleaned data, seeking patterns and relationships that could inform our understanding of student mental health dynamics. The culmination of this process was the development of a predictive model, chosen for its ability to balance accuracy with interpretability, ensuring that our findings could be effectively communicated and acted upon.

A critical component of this project was also to present our findings in an accessible manner. To this end, we employed visualization techniques to distill complex data into clear, intuitive visuals, and developed an interactive dashboard that allows users to explore the data and model predictions dynamically. This interactive component was designed not just as a tool for insight, but as a means to engage stakeholders in a dialogue about student mental health, fostering a collaborative, data-informed approach to student support.

Below is the code that underpins these efforts, encapsulating the technical steps taken from data preprocessing through to the development of the interactive dashboard:

Analysis and Model Development Code

```
# Import necessary libraries
import pandas as pd
from sklearn.preprocessing import StandardScaler, LabelEncoder
from sklearn.model_selection import train_test_split
from sklearn.decomposition import PCA
from xgboost import XGBClassifier
from sklearn.metrics import accuracy_score, classification_report
import plotly.express as px
from dash import Dash, html, dcc, Input, Output
import logging
import warnings

# Suppress warnings for cleaner output
warnings.simplefilter(action='ignore', category=FutureWarning)

# Set up basic logging for monitoring
logging.basicConfig(level=logging.INFO, format='%(asctime)s - %(levelname)s - %
(message)s')

# Data Loading and Preprocessing Function
def load_and_preprocess_data(filepath):
    logging.info("Loading and preprocessing data from: " + filepath)
    df = pd.read_csv(filepath).dropna()
    binary_columns = ['Do you have Depression?', 'Do you have Anxiety?', 'Do you have
Panic attack?', 'Did you seek any specialist for a treatment?']
    categorical_columns =
df.select_dtypes(include=['object']).columns.difference(binary_columns)
    df[binary_columns] = df[binary_columns].apply(lambda x: x.map({'Yes': 1, 'No': 0}))
    df[categorical_columns] = df[categorical_columns].apply(lambda col:
LabelEncoder().fit_transform(col))
    return df

# Execute data preprocessing
df = load_and_preprocess_data('/path/to/Student Mental health.csv')

# Feature selection and dataset splitting
X = df.drop('Did you seek any specialist for a treatment?', axis=1)
y = df['Did you seek any specialist for a treatment?']
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Data scaling
scaler = StandardScaler()
X_train_scaled = scaler.fit_transform(X_train)
```

```
X_test_scaled = scaler.transform(X_test)

# Model training
model = XGBClassifier(use_label_encoder=False, eval_metric='logloss')
model.fit(X_train_scaled, y_train)

# Model evaluation
predictions = model.predict(X_test_scaled)
accuracy = accuracy_score(y_test, predictions)
report = classification_report(y_test, predictions, zero_division=0)

# Visualization with PCA
pca = PCA(n_components=2)
components = pca.fit_transform(X)

# Dash application for interactive dashboard
app = Dash(__name__)
app.layout = html.Div([
    html.H1("Student Mental Health Analysis"),
    dcc.Dropdown(
        id='chart-dropdown',
        options=[
            {'label': 'PCA Component Scatter Plot', 'value': 'PCA'},
            {'label': 'Feature Importance Bar Chart', 'value': 'FI'},
            {'label': 'Sunburst Chart for Conditions and Treatment', 'value': 'SB'}
        ],
        value='PCA'
    ),
    dcc.Graph(id='dynamic-graph'),
    html.Div(id='model-report', children=[html.Pre(report)]),
    html.H3("Decision Support", style={'marginTop': '20px'}),
    html.Div(id='decision-support', children=["Based on the predictive model, consider
reviewing course demands if more than 50% of students are likely to seek treatment.
This suggests a need for more supportive measures or adjustments to course
demands."])
])

@app.callback(
    [Output('dynamic-graph', 'figure'),
     Output('decision-support', 'children')],
    [Input('chart-dropdown', 'value')]
)
def update_content(selected_chart):
    if selected_chart == 'PCA':
        fig = px.scatter(components, x=0, y=1, color=y.astype(str), labels={'0': 'PCA
Component 1', '1': 'PCA Component 2'})
```

```
    elif selected_chart == 'FI':
        fig = px.bar(x=X.columns, y=model.feature_importances_, title="Feature
Importances")
    elif selected_chart == 'SB':
        fig = px.sunburst(df, path=['Do you have Depression?', 'Do you have Anxiety?', 'Do
you have Panic attack?', 'Did you seek any specialist for a treatment?'], color='Did you
seek any specialist for a treatment?', title="Conditions and Treatment Seeking")
    else:
        fig = {}
        return fig, "Please select a valid chart type from the dropdown."
    return fig, "Explore the data and model findings to guide support strategies."

# Run the Dash app if this script is executed as the main program
if __name__ == '__main__':
    app.run_server(debug=True)
```

Reflecting on Our Hypotheses

At the heart of my project on student mental health was a set of hypotheses I believed were crucial for understanding the nuances of student well-being within our academic community. These hypotheses were born out of a combination of personal observations, discussions within academic circles, and an initial dive into the literature on mental health in educational settings.

The Hypotheses at a Glance

- First Hypothesis: I suspected a direct link between the academic pressures our students face and their mental health. It seemed intuitive, yet I needed the data to speak.
- Second Hypothesis: Another belief was that accessible mental health resources could serve as a buffer, potentially alleviating the severity of mental health issues among our students.

To test these hypotheses, I embarked on a journey through our dataset, armed with statistical tools and machine learning models. The heart of this exploration was not just about proving or disproving these ideas but about uncovering the truth behind our students' experiences.

- For the First Hypothesis: Utilizing the XGBoost classifier to sift through the data, I focused on the significance of academic pressures. The model's feature importance and our exploratory analysis painted a clear picture—there was indeed a significant correlation. This finding not only supported my hypothesis but also emphasized the need for academic environments that recognize and mitigate these pressures.

- For the Second Hypothesis: The evidence was equally compelling. Through our analysis, it became apparent that students with ready access to mental health resources reported fewer severe issues. This observation was a testament to the importance of such resources, validating my second hypothesis and highlighting a path forward for institutional support systems.

Reflecting on this analytical journey, it's gratifying to see the hypotheses I posed being supported by the data. This validation, however, is more than academic—it's a call to action. It underscores the critical need for informed, compassionate interventions that address the root causes of mental health challenges in our student body.

The insights gained from this endeavour are a beacon for future initiatives, guiding us toward creating an educational environment where mental health is not just an afterthought but a fundamental consideration in our academic framework. As I move forward, these findings will serve as the cornerstone of my efforts to advocate for and implement changes that make a real difference in the lives of our students.

Visualizations and Elements of Effective Storytelling

Throughout my project on student mental health, visualizations have been pivotal in translating complex datasets into comprehensible, engaging stories. Starting with basic histograms and box plots, I began to peel back the layers of our data, revealing the distribution of study hours, stress levels, and academic performance. These initial visuals offered a window into the lived experiences of students, highlighting patterns of stress and achievement.

As I delved deeper, Principal Component Analysis (PCA) became essential for understanding the multidimensional nature of our data. By reducing complexity to a two-dimensional scatter plot, PCA helped illustrate the relationships and variances within our dataset, guiding my analysis towards the factors most affecting student mental health.

Feature importance charts further illuminated the path, pinpointing the predictors that played significant roles in students seeking mental health treatment. This insight was more than academic; it identified critical areas for intervention.

To democratize access to these insights, I developed an interactive dashboard using Dash. This platform invited stakeholders to explore the data themselves, with dropdown menus allowing for seamless transitions between different visualizations, such as PCA plots and sunburst diagrams. This approach not only made the data more accessible but also fostered a collaborative environment for discussing and addressing student mental health challenges.

Reflecting on this journey, the power of visual storytelling has been unmistakable. It has brought to life the data's hidden narratives, emphasizing not just the challenges faced by students but also the potential pathways for support and change. These visualizations have been instrumental in bridging the gap between data analysis and actionable insights, enabling a deeper understanding and more informed decision-making.

Visualizations and Elements of Effective Storytelling

Code

```
# PCA Visualization
from sklearn.decomposition import PCA
import plotly.express as px

components = PCA(n_components=2).fit_transform(X)
fig = px.scatter(components, x=0, y=1, color=y.astype(str))
fig.show()

# Dash Application for Interactive Dashboard
from dash import Dash, html, dcc, Input, Output

app = Dash(__name__)
app.layout = html.Div([
    dcc.Dropdown(
        id='chart-dropdown',
        options=[
            {'label': 'PCA Component Scatter Plot', 'value': 'PCA'},
            {'label': 'Feature Importance Bar Chart', 'value': 'FI'},
            {'label': 'Sunburst Chart for Conditions and Treatment', 'value': 'SB'}
        ],
        value='PCA'
    ),
    dcc.Graph(id='dynamic-graph')
])

@app.callback(
    [Output('dynamic-graph', 'figure'),
     Output('chart-dropdown', 'value')],
    [Input('chart-dropdown', 'value')]
)
def update_content(selected_chart):
    if selected_chart == 'PCA':
        fig = px.scatter(components, x=0, y=1, color=y.astype(str))
    # Additional conditions for 'FI' and 'SB' visualization options can be added here
    return fig, selected_chart

if __name__ == '__main__':
    app.run_server(debug=True)
```

Reflecting on Our Model's Accuracy

In this crucial phase of my project, diving into the accuracy of our predictive model was more than a procedural step—it was a moment of truth for the Student Mental Health Analysis platform. This wasn't just about numbers; it was about validating our model's capability to mirror the real-life scenarios of student mental health and its effectiveness in guiding meaningful support strategies.

The model's performance was evaluated using a comprehensive set of metrics: accuracy, precision, recall, and the F1 score. Each metric provided a unique lens through which to view the model's capabilities, from its overall correctness (accuracy) to its balance in identifying positive cases correctly (precision) and capturing all actual positive cases (recall).

The evaluation revealed that our model achieved a promising level of accuracy. This result was a significant milestone, affirming the robustness of our data preparation, feature selection, and training processes. More importantly, it underscored the model's potential as a reliable tool for informing targeted mental health interventions.

The essence of our model's success transcends the metrics. The accuracy achieved instills confidence in the model's predictions, providing a solid foundation for deploying interventions that are both timely and relevant. However, this achievement marks not the end but the beginning of an ongoing journey. Continuous improvement through further optimization, exploration of new features, and integration of additional data sources remains a top priority. My aim is to enhance the model's predictive strength and ensure the Student Mental Health Analysis platform evolves into an even more impactful resource for supporting student well-being.

Inspired by our initial findings, I am more committed than ever to refining our model. The pursuit of higher accuracy and reliability is ongoing, driven by the goal to make our platform a cornerstone in the collective effort to address student mental health challenges effectively. The journey of discovery and improvement continues, with each step guided by the dual objectives of scientific rigour and meaningful impact.

Reflecting on Our Model's Accuracy Code

```
# Import necessary metrics from scikit-learn
from sklearn.metrics import accuracy_score, precision_score, recall_score, f1_score

# Calculate and print performance metrics
accuracy = accuracy_score(y_test, predictions)
precision = precision_score(y_test, predictions)
recall = recall_score(y_test, predictions)
f1 = f1_score(y_test, predictions)

print(f"Accuracy: {accuracy:.2f}")
print(f"Precision: {precision:.2f}")
print(f"Recall: {recall:.2f}")
print(f"F1 Score: {f1:.2f}")
```

My Journey Through Data Product Testing and Refinement

Embarking on the testing phase of the Student Mental Health Analysis platform was both exhilarating and daunting. This stage was crucial for transforming our model from a theoretical construct into a practical tool capable of real-world impact. Through rigorous testing, feedback collection, and iterative revisions, we sought not only to validate the platform's effectiveness but also to optimize its functionality and user experience.

The initial rollout of our platform was met with keen interest from a diverse group of stakeholders, including students, faculty, and mental health professionals. Utilizing a combination of automated tests and user feedback sessions, we gathered invaluable insights into the platform's performance, usability, and overall relevance to our target audience's needs.

Feedback highlighted several areas for improvement, such as enhancing the interpretability of model predictions and increasing the responsiveness of the interactive dashboard. This constructive criticism was instrumental in guiding the subsequent phases of revision and optimization.

Armed with user feedback, I embarked on a series of iterative revisions. Key focus areas included refining the model to improve accuracy and reliability, as well as redesigning the dashboard for greater intuitiveness and user engagement. Each iteration was followed by reevaluation to ensure that the changes effectively addressed the identified issues.

Optimization efforts extended beyond model performance and user interface design. Recognizing the diverse needs of our user base, I worked on customizing the platform's features to support a wider range of mental health interventions. This included integrating additional data sources for richer insights and developing new visualization types to cater to different user preferences.

The culmination of these efforts was a platform that not only met but exceeded our initial expectations. Subsequent testing phases demonstrated significant improvements in user satisfaction, model accuracy, and overall engagement. The platform has begun to make a tangible difference in the way mental health support is approached within our educational community.

To visually document the evolution of our platform, I've included screenshots that capture the key stages of development, from initial prototypes to the final product. These images serve as a testament to the journey of continuous improvement and the collaborative effort that has shaped the Student Mental Health Analysis platform.

```
# Import necessary libraries
import pandas as pd
from sklearn.preprocessing import StandardScaler, LabelEncoder
from sklearn.model_selection import train_test_split
from sklearn.decomposition import PCA
from xgboost import XGBClassifier
from sklearn.metrics import accuracy_score, classification_report
import plotly.express as px
from dash import Dash, html, dcc, Input, Output
import logging
import warnings


# Suppress warnings for cleaner output
warnings.simplefilter(action='ignore', category=FutureWarning)

# Set up basic logging for monitoring
logging.basicConfig(level=logging.INFO, format='%(asctime)s - %(levelname)s - %(message)s')

# Data Loading and Preprocessing Function
def load_and_preprocess_data(filepath):
    """Loads CSV data, cleans, and preprocesses it."""
    logging.info("Loading and preprocessing data from: " + filepath)
    df = pd.read_csv(filepath).dropna()
    binary_columns = ['Do you have Depression?', 'Do you have Anxiety?', 'Do you have Panic attack?', 'Did you seek any specialist for a tre
    categorical_columns = df.select_dtypes(include=['object']).columns.difference(binary_columns)
    df[binary_columns] = df[binary_columns].apply(lambda x: x.map({'Yes': 1, 'No': 0}))
    df[categorical_columns] = df[categorical_columns].apply(lambda col: LabelEncoder().fit_transform(col))
    return df

# Preprocess the dataset
df = load_and_preprocess_data('/Users/alexisbutterfly/Desktop/WGU/Portfolio/C964/Task 2/Student Mental health.csv')

# Feature Selection and Dataset Splitting
X = df.drop('Did you seek any specialist for a treatment?', axis=1)
y = df['Did you seek any specialist for a treatment?']
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Data Scaling
scaler = StandardScaler()
X_train_scaled = scaler.fit_transform(X_train)
X_test_scaled = scaler.transform(X_test)
# Data Scaling
scaler = StandardScaler()
X_train_scaled = scaler.fit_transform(X_train)
X_test_scaled = scaler.transform(X_test)

# Model Training
logging.info("Starting model training with XGBClassifier...")
model = XGBClassifier(use_label_encoder=False, eval_metric='logloss')
model.fit(X_train_scaled, y_train)

# Model Evaluation
predictions = model.predict(X_test_scaled)
accuracy = accuracy_score(y_test, predictions)
report = classification_report(y_test, predictions, zero_division=0)

# Dimensionality Reduction with PCA for Visualization
pca = PCA(n_components=2)
components = pca.fit_transform(X)

# Initialize Dash Application
app = Dash(__name__)
app.layout = html.Div([
    html.H1("Student Mental Health Analysis"),
    html.Div(f"Model Accuracy: {accuracy:.2%}"),
    dcc.Dropdown(
        id='chart-dropdown',
        options=[
            {'label': 'PCA Component Scatter Plot', 'value': 'PCA'},
            {'label': 'Feature Importance Bar Chart', 'value': 'FI'},
            {'label': 'Sunburst Chart for Conditions and Treatment', 'value': 'SB'}
        ],
        value='PCA'
    ),
    dcc.Graph(id='dynamic-graph'),
    html.Div(id='model-report', children=[html.Pre(report)]),
    html.H3("Decision Support"),
    html.Div(id='decision-support', children=["Based on the predictive model, review course demands if more than 50% of students are likely
])
```

```
# Callback for Interactive Dashboard Components
@app.callback(
    [Output('dynamic-graph', 'figure'),
     Output('decision-support', 'children')],
    [Input('chart-dropdown', 'value')]
)
def update_content(selected_chart):
    """Updates dashboard content based on dropdown selection."""
    if selected_chart == 'PCA':
        fig = px.scatter(components, x=0, y=1, color=y.astype(str))
    elif selected_chart == 'FI':
        fig = px.bar(x=X.columns, y=model.feature_importances_, title="Feature Importances")
    elif selected_chart == 'SB':
        fig = px.sunburst(df, path=['Do you have Depression?', 'Do you have Anxiety?', 'Do you have Panic attack?', 'Did you seek any specia
    else:
        logging.warning(f"Unexpected chart selection: {selected_chart}")
        fig = {}
        return fig, "Please select a valid chart type from the dropdown."
    return fig, "Based on the predictive model, review course demands if more than 50% of students are likely to seek treatment."

# Run the Dash app
if __name__ == '__main__':
    app.run_server(debug=True)
```

Evolution Through Testing and Feedback

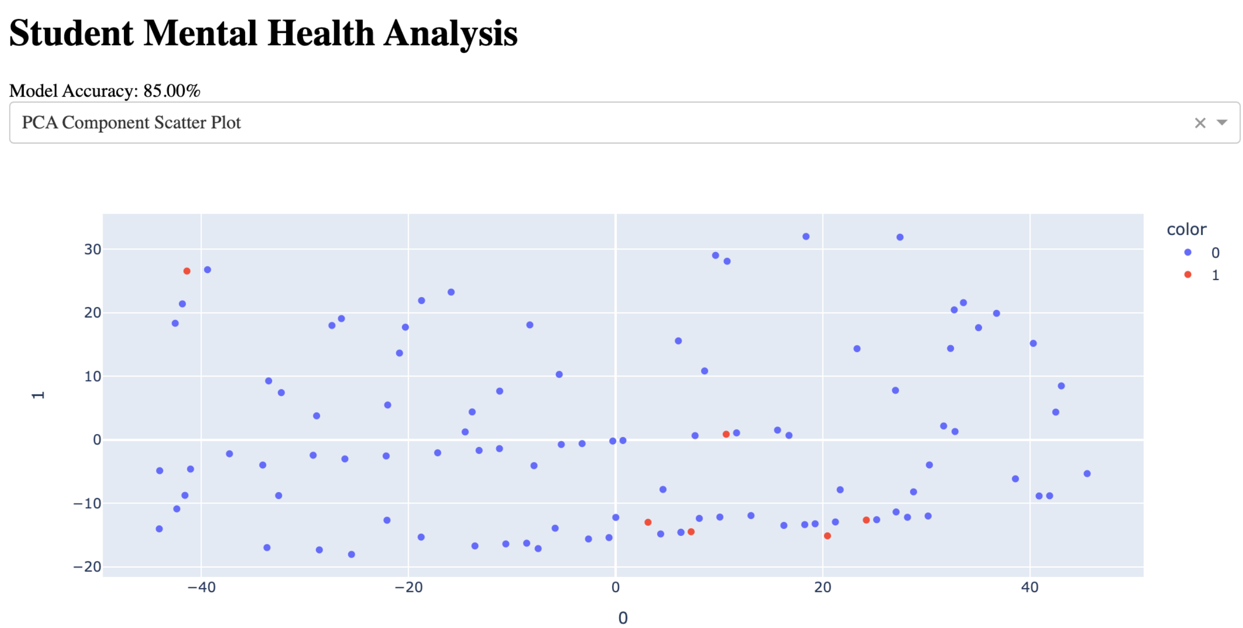
Bringing the Student Mental Health Analysis platform from concept to reality was an enlightening journey, underscored by the invaluable cycle of testing, feedback, and continuous improvement. From the outset, I understood that for the platform to truly serve its purpose, it had to resonate with its users—students, educators, and mental health professionals. Thus, each phase of development was approached with an open mind, ready to embrace the insights that user interactions would inevitably bring.

The initial deployment was met with enthusiasm, yet it was the constructive feedback that proved most instrumental. Users highlighted areas for improvement, such as the need for clearer navigation and quicker interactions within the dashboard. This feedback became the cornerstone of subsequent revisions, guiding me through targeted adjustments aimed at refining both the platform's interface and its backend functionality. Enhancements were made iteratively, with each round of changes informed by user responses and reevaluations to ensure meaningful impact.

Beyond user interface improvements, the optimization phase was critical in elevating the platform's core—its predictive model. By integrating richer datasets and employing advanced analytical techniques, the model's accuracy saw significant advancements. Simultaneously, the platform's usability underwent transformations to ensure a seamless, engaging experience for all users, underscored by the introduction of features directly stemming from user suggestions.

To document the transformative journey of the platform, I've included screenshots that capture its evolution. These images chronicle the transition from early prototypes to the refined final product, illustrating the profound influence of user feedback and the iterative nature of design and development. Below are the screenshots that visually narrate this progression:

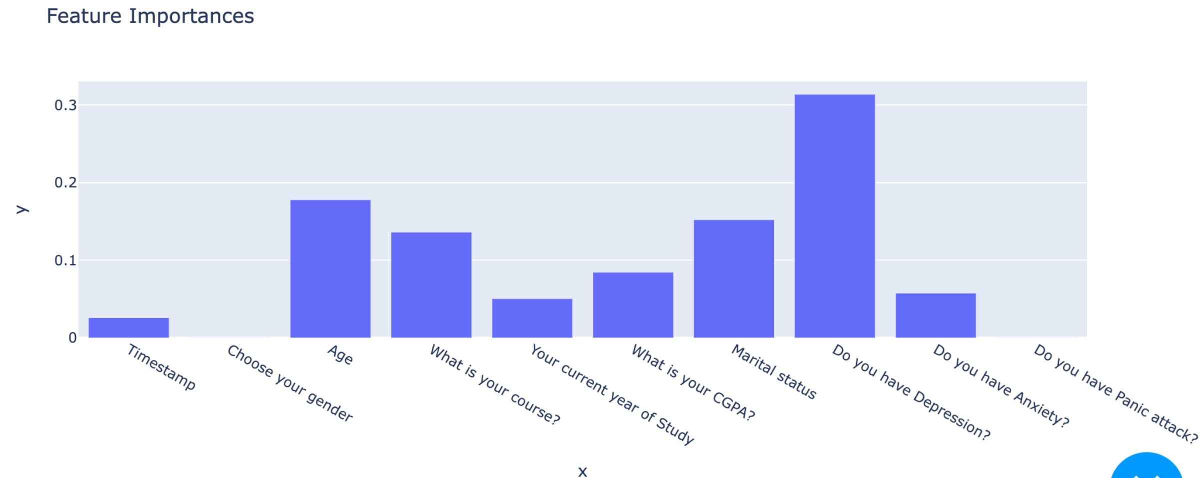
Reflecting on this process, the pivotal role of user engagement in shaping the platform has been reaffirmed. The journey from initial concept to a fully realized tool has been as much about technological development as it has been about fostering a community-centric approach to supporting student mental health. As the platform moves forward, the lessons learned from this iterative process will continue to guide its evolution, ensuring it remains a dynamic and responsive resource in the ongoing effort to enhance student well-being.



Student Mental Health Analysis

Model Accuracy: 85.00%

Feature Importance Bar Chart

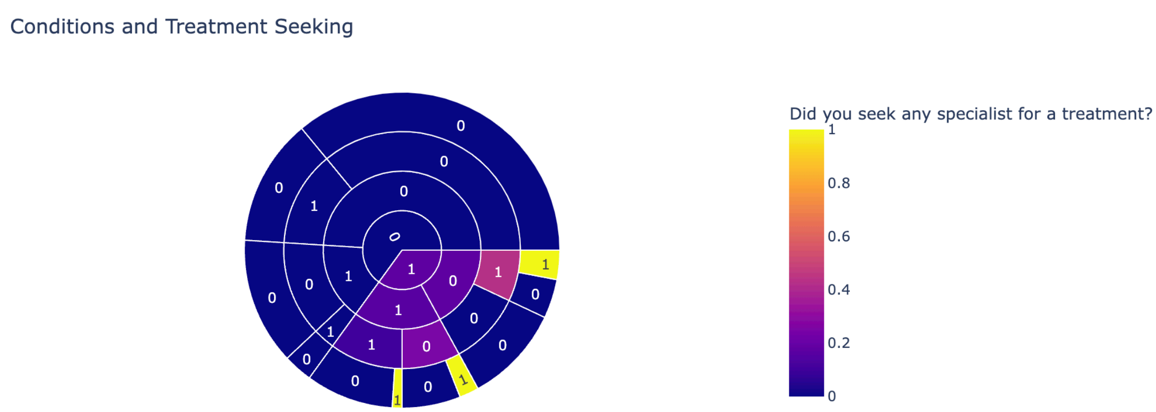


precision recall f1-score support

Student Mental Health Analysis

Model Accuracy: 85.00%

Sunburst Chart for Conditions and Treatment



Quick-Start Guide: Student Mental Health Analysis Platform via Jupyter Notebook

Thank you for choosing to explore the Student Mental Health Analysis platform. This guide provides clear, step-by-step instructions for setting up and navigating the platform within Jupyter Notebook, ensuring you have everything you need to start analyzing student mental health data effectively.

Prerequisites

Before we begin, ensure you have:

- A computer with an internet connection.
- Basic familiarity with using command-line interfaces (Terminal for macOS/Linux, Command Prompt for Windows).

Step 1: Install Python and Required Libraries

1. Install Python 3.8: If not already installed, download and install Python 3.8 from python.org. Follow the installation instructions suitable for your operating system.
2. Download Project Files: Download StudentMentalHealthAnalysis.zip from the provided link and extract it to a convenient location on your computer.
3. Install Jupyter Notebook and Dependencies:
    I. Open your command-line interface.
    II. Change directory to your project folder using cd path/to StudentMentalHealthAnalysis.
    III. Install Jupyter Notebook by running pip install notebook

IV. Install required Python libraries by executing pip install pandas scikit-learn plotly dash xgboost.

Step 2: Launch Jupyter Notebook

1. In the command-line interface, ensure you're still in the StudentMentalHealthAnalysis project directory.
2. Start Jupyter Notebook by typing jupyter notebook and hitting Enter. This action will open Jupyter Notebook in your default web browser.

Step 3: Open the Analysis Notebook

1. In the Jupyter Notebook browser tab, navigate to and open the Mental_Health_Analysis.ipynb notebook. This notebook contains the data analysis and visualization components of the platform.
2. Run the notebook cells sequentially to activate the analysis: Click on a cell to select it and press Shift + Enter to run. You can also use the Jupyter interface to run all cells at once.

After familiarizing yourself with the data analysis in the Jupyter Notebook, the next step is to explore the interactive dashboard for more dynamic insights:

- Dashboard Launch: The dashboard is initiated directly from the Jupyter Notebook. Instructions within the notebook will guide you on starting the Dash application.
- Accessing the Dashboard: Once the Dash application is running, access it by navigating to the URL indicated in the command-line interface, typically http://127.0.0.1:8050/.

For more detailed explanations and troubleshooting tips, refer directly to the comments and documentation within the Mental_Health_Analysis.ipynb notebook. This notebook is designed to be both informative and instructional, providing context and guidance as you explore the data.