Comparative Study of Machine Learning Methods: Predicting Criminal Justice Involvement and Mental Health Treatment Intensity

Nikolaos Gkmpenompa x22213708

Higher Diploma in Science in Data Analytics

Abstract – Mental health services must balance increasing demands for care with limited resources while addressing social implications such as criminal justice involvement. Predictive analysis offers a potential approach to identify individuals at higher risk of negative influences and to optimise service delivery. This study applies and compares five machine learning methods - Logistic Regression, Naive Bayes, Decision Tree, K-Nearest Neighbors (KNN), and Random Forest - to two related datasets from the Office of Mental Health (OMH) Patient Characteristics Survey (PCS). Using data from 2019 and 2022, this study investigates methods to (1) predict whether a mental health patient will require intensive treatment (inpatient, residential, or emergency) versus non-intensive treatment (outpatient or support services), and (2) identify patient characteristics associated with criminal justice involvement. Both datasets contain demographic, socioeconomic, and clinical data from over 20,000 patients each. Through data preprocessing, exploratory data analysis, and model evaluation, this study presents significant findings on the performance and application of different ML techniques in mental health conditions. Results indicate that Naïve Bayes achieved the highest predictive performance for both tasks, with Random Forest and Logistic Regression displaying great universality across datasets.

Keywords— Machine learning, Predictive modeling, Mental health services, Supervised learning, Logistic Regression, Naïve Bayes, Decision Tree, KNN, Random Forest

I. INTRODUCTION

Optimised planning of mental health treatment and risk assessment are critical applications of predictive analysis in healthcare. Recent advances in machine learning have shown promising results in predicting mental health crises from electronic health records, with models achieving AUC scores of 0.797 for continuous patient monitoring [1]. Intensive treatments, such as inpatient or emergency care, are costly and demand resources, making accurate identification of such patients critical for efficient service delivery. Additionally, given the proven connection between neglected mental health problems and encounters with the justice system, understanding the factors associated with criminal justice involvement can support early interventions and reduce social and economic consequences. The motivation for choosing this topic is my interest in sociological subjects touching human behavior and by extension mental health.

This study addresses two related research questions:

1. Which factors predict whether a mental health patient will require intensive or non-intensive treatment?

2. Which factors are most predictive of criminal justice involvement among mental health patients?

To answer these questions, five machine learning methods are utilised: Logistic Regression, Naïve Bayes, Decision Tree, k-Nearest Neighbors, and Random Forest. Each model is trained and tested on pre-processed PCS data. Their performance is compared using multiple evaluation metrics. By analysing the model's performance across two related datasets with different prediction targets, this study aims to identify effective approaches, understand their limitations, and highlight key predictors relevant to both treatment planning and legal issues prevention.

II. RELATED WORK

Research has shown that individuals with serious mental illness have disproportionately high rates of criminal legal system involvement, highlighting the critical need for predictive interventions [2]. Furthermore, studies examining mental health treatment patterns among justice-involved individuals have revealed significant disparities in access to care and treatment outcomes [3]. Using a comparison approach, this study examines the various benefits and limitations of five different machine learning methods in mental health prediction applications. Using PCS datasets from two different periods, 2019 and 2022, collected by the New York State Office of Mental Health (OMH). This research examines how different factors affect model performance. The PCS is conducted every two years, and collects demographic, clinical, and service-related information for each person who receives a public mental health service during a specified one-week period [4]. The New York State Office of Mental Health (OMH) oversees a large multi-faceted mental health system serving nearly 800,000 individuals each year. This includes operating 23 psychiatric centers and overseeing the Nathan S. Kline Institute and New York Psychiatric Institute. OMH regulates and certifies more than 6,500 programs operated by local governments and nonprofit agencies [5].

A. A. Mental Health Prediction using Machine Learning

Recent studies demonstrate increasing interest in machine learning application in mental health prediction. Fernandes et al. [6] successfully applied ensemble methods to predict depression severity, finding that Random Forest outperformed traditional statistical methods. Similarly, Chekroud et al. [7] used machine learning to predict antidepressant treatment results, highlighting the superior performance of non-linear algorithms over traditional methods. Regarding the prediction of mental health treatment intensity, studies has shown interesting results. Burke et al. [8] compared multiple algorithms to predict psychiatric hospitalisation, finding that tree-based methods performed well specifically with electronic health record data. Their work highlights the importance of feature engineering in healthcare applications. Berk et al. [9] used machine learning to predict criminal behavior while considering mental health issues. They demonstrated the difficulty of involving psychological variables into predictive models.

III. METHODOLOGY

The methodology follows a standard machine learning strategy: data preprocessing and cleaning, exploratory data analysis, categorical variable encoding, feature selection, model development, and performance evaluation. Supervised machine learning techniques are employed to develop binary classification models for two distinct mental health prediction tasks. The first model predicts treatment intensity requirements using the 2019 PCS data and the second one, criminal justice involvement status using the 2022 PCS data. Sampling techniques were implemented to create balanced datasets suitable for classification algorithms. This approach uses demographic, socioeconomic, and clinical factors to build predictive models that can assist healthcare services in treatment planning and resource allocation.

B. Data Source

The two datasets that are utilised are publicly available at the Office of Mental Health Patient Characteristics Survey:

Treatment Intensity Prediction (2019 PCS): Contains demographic, socioeconomic, mental illness, and substance use variables from individuals receiving mental health services. The target variable is separated between intensive treatment requirements (inpatient/residential/emergency) and non-intensive treatment (outpatient/support services).

Criminal Justice Involvement Prediction (2022 PCS): Includes similar demographic, socioeconomic, mental illness, and substance use variables, with the target variable distinguished between patients with criminal justice involvement and without criminal justice involvement.

Both datasets from New York State's public mental health system contain real data from anonymised patients, ensuring compliance with privacy and confidentiality standards. They both fulfil the project requirements with over 20,000 rows each and more than 10 columns.

C. Data Preparation

For this project, Jupiter Notebook was utilised and pandas library for dataframe management.Following a careful initial inspection, the following variables were selected for both datasets: 'Age Group', 'Sex', 'Race', 'Education Status', 'Employment Status', 'Living Situation', 'Region Served', 'Mental Illness', 'Serious Mental Illness', 'Alcohol Related Disorder', 'Drug Substance Disorder', and 'Program Category'. The variables 'No Insurance' and 'Opioid Related Disorder' were used only in (2019 PCS) dataset, and the variables 'Cannabis Recreational Use' and 'Criminal Justice Status' were used only in (2022 PCS) dataset.

D. Data Cleaning

Due to the datasets initial size of almost 200,000 records, much larger than the project's minimum requirement of 10,000 rows, the records with missing socioeconomic and demographic data were aggressively filtered, prioritising accuracy and consistency over preserving the greatest number of original entries. Following inspection of the variable instances in both datasets, first, all instances of 'UNKNOWN', 'UNKNOWN RACE', and 'UNKNOWN EMPLOYMENT STATUS' were replaced with NaN values, to standardise missing data representation.

```
# Replacing 'UNKNOWN' values with NaN in the entire dataset
ML_CA2_Criminal = ML_CA2_Criminal.replace(['UNKNOWN','UNKNOWN RACE','UNKNOWN EMPLOYMENT STATUS'], np.nan)

# Dropping rows with any NaN values
ML_CA2_Criminal = ML_CA2_Criminal.dropna()
```

```
# Replacing 'UNKNOWN' values with NaN in the entire dataset
ML_CA2_Project = ML_CA2_Project.replace(['UNKNOWN','UNKNOWN RACE','UNKNOWN EMPLOYMENT STATUS'], np.nan)

# Dropping rows with any NaN values
ML_CA2_Project = ML_CA2_Project.dropna()
```

Any rows containing NaN values were dropped from the dataset using the dropna() function.

E. Data Grouping and Standardisation

For easier analysis and to reduce dimensionality, the datasets were standardised and grouped into categories. In both datasets, in the Race column, 'MULTI-RACIAL' and 'OTHER' were combined into a single 'OTHER/MULTI-RACIAL' category. The Education Status column was simplified by merging 'NO FORMAL EDUCATION' and 'OTHER' into one 'OTHER' category. To improve readability and consistency, long names across multiple variables were shortened using mapping dictionaries. Education levels were simplified (e.g., 'MIDDLE SCHOOL TO HIGH SCHOOL' became 'HIGH SCHOOL'), employment statuses were shortened (e.g., 'NOT IN LABOR FORCE: UNEMPLOYED AND NOT LOOKING FOR WORK' became 'NOT IN LABOR FORCE'), and regional names were abbreviated for clarity.

```
[27]: # merging MULTI-RACIAL and OTHER into one category "OTHER/MULTI-RACIAL" in Rcae column
ML_CA2_Criminal['Race'] = ML_CA2_Criminal['Race'].replace(['MULTI-RACIAL', 'OTHER'], 'OTHER/MULTI-RACIAL')

[29]: # merging NO FORMAL EDUCATION and OTHER into one category "OTHER EDUCATION" in Education Status column
ML_CA2_Criminal['Education Status'] = ML_CA2_Criminal['Education Status'].replace(['NO FORMAL EDUCATION', 'OTHER'], 'OTHER')
```

In the (2019 PCS) dataset, the target variable 'Program Category' was restructured based on treatment intensity. Low-intensity programs ('OUTPATIENT' and 'SUPPORT') were grouped as 'NON-INTENSIVE', while high-intensity programs ('RESIDENTIAL', 'INPATIENT', and 'EMERGENCY') were categorized as 'INTENSIVE'. Respectively in the (2022 PCS) dataset, the target variable 'Criminal Justice Status' instances 'NO' and 'YES' were replaced with 'NO RECORD' and 'RECORDED' for clarity.

F. Data Sampling

Treatment Intensity Prediction (2019 PCS): For the treatment intensity classification task, equal representation was reached by sampling 30,000 cases from both 'INTENSIVE' and 'NON-INTENSIVE' program categories. This created a balanced dataset of 60,000 cases with a 50:50 class distribution.

```
[49]: # importing necessary library
from sklearn.utils import resample

# separating the patients between ones requiring intensive treatment and non-intensive treatment.
intensive = ML_CA2_Project[ML_CA2_Project['Program Category'] == 'INTENSIVE']
non_intensive = ML_CA2_Project[ML_CA2_Project['Program Category'] == 'NON-INTENSIVE']

[51]: # sampling 30k random cases from each category
intensive_sample = resample(intensive, n_samples=30000, random_state=42)
non_intensive_sample = resample(non_intensive, n_samples=30000, random_state=42)

[53]: # joining both sample datasets to form the final balanced one
ML_CA2_Balanced = pd.concat([intensive_sample, non_intensive_sample])
```

The sampling used replacement where necessary and maintained the same random state (42) for consistency. Column names were similarly standardised and 'Program Category' renamed to 'treatment_intensity' as the target variable.

Criminal Justice Involvement Prediction (2022 PCS): To address class imbalance in the Criminal Justice Status variable, random sampling was performed using "resample" function. The original dataset contained 134,185 cases with 'NO' criminal justice involvement and 9,739 cases with 'YES' involvement, creating significant imbalance.

```
[37]: # importing necessary library
from sklearn.utils import resample

# separating the patients between with Criminal Justice Status = 'YES' and with 'NO'
criminal_record = ML_CA2_Criminal[ML_CA2_Criminal['Criminal Justice Status'] == 'YES']
no_criminal_record = ML_CA2_Criminal[ML_CA2_Criminal['Criminal Justice Status'] == 'NO']
```

The 'NO' category was downsampled from 134,185 to 12,000 cases, while the 'YES' category was downsampled from 9,739 to 8,000 cases. Both samples were created without replacement using a random state (42) for reproducibility.

```
[39]: # For NO cases, downsampling randomly from 134,185 to 12,000
no_criminal_sampled = resample(no_criminal_record, replace=False, n_samples=12000, random_state=42)

# For YES cases, downsampling randomly from 9,739 to 8,000
criminal_sampled = resample(criminal_record, replace=False, n_samples=8000, random_state=42)
```

```
[43]: # Combining sampled datasets
joined_criminal = pd.concat([no_criminal_sampled, criminal_sampled])

# Shuffle the combined dataset
ML_CA2_Criminal = joined_criminal.sample(frac=1, random_state=42).reset_index(drop=True)
```

Then, the sampled datasets were combined and shuffled to create the final balanced dataset of 20,000 cases. Like (2019 PCS), the column names were standardised for consistency, and the target variable 'Criminal Justice Status' became 'criminal_involvement'.Both processed datasets were exported as CSV files for analysis, named 'CA2_treatment_intensity.csv' and 'CA2_criminal_involvement.csv'.

G. *Exploratory Data Analysis*

Treatment Intensity Prediction (2019 PCS) - Key Findings

Demographic Factors: Adults required intensive treatment more frequently compared to children, 56.5% to 15.9%, and males more than females, 56.2% to 43.1%. Black patients showed the highest intensive treatment rates of 63.5%, followed by White at 46.6% and Other/Multi-racial patients at 36.9%.

Socioeconomic Factors: High school graduates had higher intensive treatment needs, 55.2% and elementary-educated patients had the lowest, 15.0%. Unemployed individuals and those not in the labor force required more intensive care than employed patients. As expected, institutional residents required intensive treatment in 93.0% of cases compared to 41.9% for private home residents.

Clinical Conditions: Mental health conditions predicted high treatment intensity: mental illness at 50.5%, serious mental illness at 51.7%. Substance-related disorders showed the strongest correlation with intensive treatment needs: alcohol disorders 70.4%, substance disorders 68.6%, and opioid disorders 66.6%.

Criminal Justice Involvement Prediction (2022 PCS) - Key Findings:

Demographic Factors: As expected, children showed much lower criminal involvement rates than adults at 10.6% versus 45.3%. Males had significantly higher criminal records at 53.2% compared to females at 22.9% and non-binary individuals at 9.1%. Black patients had the highest criminal involvement rates at 49.2%, followed by White at 37.3% and Other/Multi-racial patients at 32.3%.

Socioeconomic Factors: Elementary education patients had the lowest criminal involvement at 6.6% while high school graduates had highest rates at 46.3%. Unemployed individuals had higher criminal records at 47.1% than employed patients at 35.7%. Institutional residents showed extremely high criminal involvement at 88.9% compared to private home residents at 29.3%.

Clinical Conditions: Mental illness showed minimal association with criminal involvement at 40.0% versus 39.9%. However, substance-related conditions strongly predicted criminal records: substance disorders at 66.9% involvement, alcohol disorders at 65.8% involvement, and cannabis use at 52.2% involvement.

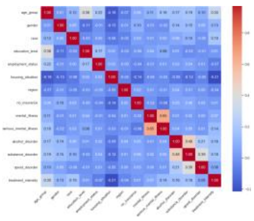
These findings give an idea of what to expect when selecting the predictors for each task in the modelling stage

IV. Modelling

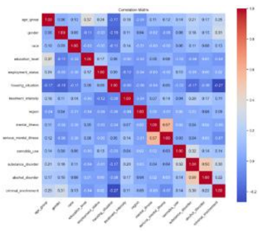
A. *Data Encoding and Feature Selection*

All categorical variables in both datasets were converted to numerical format using appropriate encoding techniques. Feature selection was performed using Pearson correlation coefficient analysis with correlation heatmaps to identify the strongest predictive variables while avoiding multicollinearity issues.

Treatment Intensity Prediction (PCS 2019): Seven features were selected based on strong correlations with the target variable: age_group (0.30), housing_situation (-0.21), race (0.19), substance_disorder (0.18), alcohol_disorder (0.16), serious_mental_illness (0.14), and gender (0.13).



Criminal Involvement Prediction (PCS 2022): The selected features were gender (0.31), substance_disorder (0.30), housing_situation (-0.27), age_group (0.25), alcohol_disorder (0.22), treatment_intensity (0.18), and education_level (varying correlations 0.04-0.37).



The feature set in both datasets demonstrated good predictive power with minimal multicollinearity.

B. *Model Selection*

The machine learning models used in this study are: Logistic Regression, Naïve Bayes, Decision Tree, K-Nearest

Neighbor (KNN) and Random Forest. These five machine learning algorithms were carefully selected to provide thorough evaluation across various modelling assumptions. The selected models offer an ideal balance between linear and non-linear approaches, parametric and non-parametric methods, individual and ensemble techniques. This diversity ensures a strong performance comparison while utilising the specific features of each dataset which contain mostly categorical features with complex interdependencies.

Logistic Regression offers highly interpretable results that are important in healthcare decision-making. Naive Bayes performs well even with small training samples and thrives with categorical data. Decision tree's rules-based interpretations against clinical that healthcare professionals can easily understand and validate. KNN provides flexibility by making no assumptions about uncovered data distributions, making it ideal for complex, non-linear relationships in patient data. Finally, Random Forest combines the interpretability advantage of decision trees with better predictive performance through ensemble training and lowers the risk of overfitting.

C. Model Training

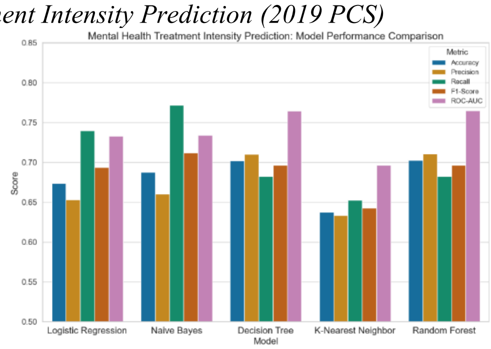
Treatment Intensity Prediction (2019 PCS): The treatment intensity prediction model utilised seven selected features: age_group, race, gender, housing_situation, serious_mental_illness, substance_disorder, and alcohol_disorder. The balanced dataset of 60,000 samples was split at an 80/20 train-test ratio with stratification, to maintain equal class distribution across both training and testing sets. A fixed random state of 42 ensured reproducible results across all models.

Criminal Justice Involvement Prediction (2022 PCS): The criminal justice involvement prediction model used seven features sorted by correlation strength: gender, substance_disorder, housing_situation, age_group, alcohol_disorder, treatment_intensity, and education_level. The imbalanced dataset of 20,000 samples was divided using an 80/20 train-test split with stratification to preserve equal distribution of criminal involvement cases. Same as the previous dataset, a random state of 42 was applied to ensure reproducible results across all models implementations.

V. Evaluation

A. Performance metrics

Treatment Intensity Prediction (2019 PCS)



Model	Accuracy	Precision	Recall	F1	ROC-AUC	FN	FP
Logistic Regression	0.67	0.65	0.74	0.69	0.73	1562	2355
Naive Bayes	0.69	0.66	0.77	0.71	0.73	1368	2379
Decision Tree	0.70	0.71	0.68	0.70	0.76	1904	1669
KNN	0.64	0.63	0.65	0.64	0.70	2085	2262
Random Forest	0.70	0.71	0.68	0.70	0.76	1903	1667

All five models showed moderate to good predictive performance, with accuracy scores ranging from 64% to 70%. Decision Tree and Random Forest emerged as the top performers, both achieving 70% accuracy, 0.70 F1-score, and 0.76 ROC-AUC. These ensemble and tree-based methods showed superior ability to capture non-linear relationships in the healthcare data provided. Naive Bayes performed well with 69% accuracy and the highest recall of 0.77, making it effective at identifying patients requiring intensive treatment. Logistic Regression achieved 67% accuracy with balanced precision-recall trade-offs. K-Nearest Neighbors showed the lowest performance at 64% accuracy, suggesting that local similarity patterns were less informative for this prediction task. Naive Bayes performed best for minimizing False Negatives (1368), making it the safest choice for ensuring patients requiring intensive treatment are not missed. Decision Tree and Random Forest performed better at minimising False Positives (1669 and 1667 respectively), making them most efficient for resource allocation by avoiding unnecessary intensive treatment staff allocation.

Criminal Justice Involvement Prediction (2022 PCS)

Model	Accuracy	Precision	Recall	F1	ROC-AUC	FN	FP
Logistic Regression	0.72	0.70	0.55	0.61	0.77	726	380
Naive Bayes	0.72	0.64	0.69	0.66	0.77	497	628
Decision Tree	0.73	0.72	0.54	0.61	0.80	741	336
KNN	0.71	0.63	0.66	0.65	0.78	539	612
Random Forest	0.73	0.72	0.54	0.62	0.80	738	340

All five models achieved strong predictive performance, with accuracy scores ranging from 71% to 73%. Decision Tree and Random Forest rose as the top performers, both achieving 73% accuracy and 0.80 ROC-AUC, demonstrating strong ability to identify complex patterns in patient's criminal justice involvement. These tree-based methods showed high precision of 0.72 but moderate recall of 0.54, indicating moderate prediction behaviour with less false positives. Logistic Regression and Naive Bayes both achieved 72% accuracy with 0.77 ROC-AUC. Naive Bayes showed the highest recall at 0.69 and best F1-score at 0.66, making it the most effective at identifying individuals with criminal involvement. K-Nearest Neighbors achieved 71% accuracy with balanced precision-recall performance, though slightly lower than other methods. Decision Tree achieved the best balance with the lowest False Positives (336), minimising inappropriate discrimination while maintaining reasonable identification of high-risk patients. Naive Bayes performed the lowest 497 False Negatives, making it most effective at identifying patients with criminal justice involvement who may profit from special care services.

B. Cross-validation

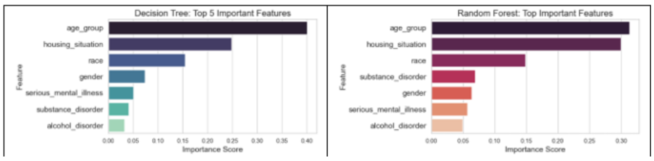
Cross-validation was only performed on KNN model in Treatment Intensity (2019 PCS) dataset. The kparameter was adjusted for better results with testing values of 5, 7, and 10.

k = 5	k = 7	k = 10
Accuracy: 0.61 Precision: 0.60 Recall: 0.66 F1-Score: 0.63 ROC-AUC: 0.69	Accuracy: 0.68 Precision: 0.73 Recall: 0.58 F1-Score: 0.65 ROC-AUC: 0.72	Accuracy: 0.64 Precision: 0.63 Recall: 0.65 F1-Score: 0.64 ROC-AUC: 0.70

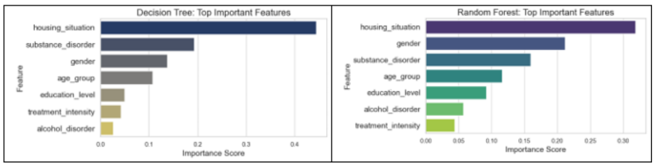
The results revealed interesting performance differences: k=5 achieved 61% accuracy with the highest recall of 0.66, k=7 reached 68% accuracy with the highest precision at 0.73, and k=10 provided balanced performance at 64% accuracy. From the mental health service perspective, the choice of k value shifts between clinical safety and resource allocation efficiency. k=5 maximises recall (0.66), ensuring that less patients needing intensive treatment are missed, aligning with clinical priorities where false negatives can lead to insufficient care towards patients. However, this approach results in lower precision (0.60), meaning more patients may be allocated to intensive treatment programs when there is no need. k=7 offers the highest precision (0.73), minimising unnecessary intensive treatment allocation and better use of resources. This approach reduces costs and prevents over-treatment but risks missing patients who need intensive care (lower recall at 0.58). k=10 provides a fair balance with moderate performance across all metrics.

C. Results comparison

Treatment Intensity Prediction (2019 PCS) - Feature Importance Findings: Both Decision Tree and Random Forest models identified age_group as the most important predictor at 40% and 31% importance respectively, followed by housing_situation (25% and 30% importance). This aligns with the exploratory analysis showing that adults and patients in institutional settings have significantly higher intensive treatment requirements. The consistent feature rankings between tree-based models validate the reliability of these predictive factors in determining treatment intensity needs. False Negatives represent patients requiring intensive treatment who are incorrectly classified as non-intensive, potentially leading to inadequate care and clinical deterioration. False Positives indicate patients receiving unnecessary intensive treatment, resulting in resource waste but maintaining patient safety through over-treatment rather than under-treatment.



Criminal Justice Involvement Prediction (2022 PCS)- Feature Importance Findings: Both Decision Tree and Random Forest models identified housing_situation as the most critical predictor, with 44% and 32% importance respectively, reflecting the strong association between institutional housing and criminal involvement found earlier in the exploratory analysis. Gender ranked as the second most important feature in Random Forest at 21% and third in Decision Tree at 14%, similar to the observed higher criminal involvement rates among males. Substance_disorder showed high importance across both models at 19% and 16%, validating its strong correlation with criminal justice involvement. The accuracy of feature rankings between tree-based models confirm the reliability of these predictive factors in identifying criminal involvement risk. False Negatives represent patients with criminal justice involvement who are not identified, potentially missing important risk factors and special care considerations in treatment planning. False Positives indicate patients incorrectly flagged for criminal involvement, which may lead to discrimination but helps in preventive support services and improved monitoring.



VI. CONCLUSION

A. Summary

This study successfully developed and evaluated machine learning models for two healthcare prediction tasks using Patient Characteristics Survey real data.
For treatment intensity prediction using the 2019 PCS dataset, Decision Tree and Random Forest were the top performers, both achieving 70% accuracy and 0.76 ROC-AUC. These tree-based methods showed strong ability to detect non-linear relationships in healthcare data while maintaining good interpretability. Naive Bayes showed the highest recall at 0.77, making it most effective for clinical safety by minimising missed patients requiring intensive treatment.

For criminal justice involvement prediction using the 2022 PCS dataset, Decision Tree and Random Forest again achieved the highest performance with 73% accuracy and 0.80 ROC-AUC. Naive Bayes again demonstrated the best recall at 0.69, being the most effective model at identifying patients with criminal justice backgrounds. Feature importance analysis suggested housing_situation, age_group, and gender as the most important predictors across both datasets, confirming the exploratory analysis findings.

B. Real Life Implications

These predictive models have strong potential to improve the distribution of resources and the delivery of mental health services. Healthcare services can utilise the treatment intensity model to enhance patient screening and to ensure appropriate care level assignments, especially in departments with limited resources. The criminal justice involvement model enables healthcare providers to identify patients who may benefit from specialised support services.

Moreover, the feature importance findings offer practical knowledge to healthcare managers. The strong predictive power of housing_situation suggests that accommodation stability should be a priority in treatment planning and resource allocation decisions. Additionally, the importance of substance-related issues in both models highlights the need of integrated mental health and addiction services.

C. Limitations

This study has several limitations that need to be mentioned. The binary classification approach is practical for decision-making but may oversimplify the complexity of treatment needs and criminal justice involvement. It's possible that significant minority cases that might have offered valuable information were removed by sampling techniques used to correct class imbalance.

D. Future Work

Future studies should explore more advanced machine learning models such as ensemble methods and gradient boosting algorithms to improve predictive accuracy. Moreover, Feature engineering approaches could include more clinical data such as medication history or time of receiving treatment. Additionally, more advanced machine learning models should be employed.

References

[1] J. Downs et al., "Machine learning model to predict mental health crises from electronic health records," Nature Medicine, vol. 28, no. 5, pp. 1240-1248, May 2022.
[2] P. Chaimowitz et al., "Forging new paths in the development of community mental health interventions for people with mental illness at risk of criminal legal system contact," Health & Justice, vol. 13, no. 1, Jan. 2025.
[3] A. L. Pinals et al., "Mental Health Treatment Among Individuals Involved in the Criminal Justice System After Implementation of the Affordable Care Act," Psychiatric Services, vol. 71, no. 8, pp. 765-773, 2020.
[4] *Patient Characteristics Survey*. (n.d.). https://omh.ny.gov/omhweb/pcs/submissions/
[5] Office of Mental Health - *About us*. (n.d.). https://omh.ny.gov/omhweb/about/
[6] A. C. Fernandes et al., "Development and evaluation of a de-identification procedure for a case register sourced from mental health electronic records," BMC Medical Informatics and Decision Making, vol. 13, no. 1, pp. 71, 2017.
[7] A. M. Chekroud et al., "Cross-trial prediction of treatment outcome in depression: A machine learning approach," The Lancet Psychiatry, vol. 3, no. 3, pp. 243-250, 2016.
[8] T. Burke et al., "Machine learning approaches to predicting psychiatric hospitalization," Journal of Medical Internet Research, vol. 21, no. 4, e12346, 2019.
[9] R. Berk, H. Heidari, S. Jabbari, M. Kearns, and A. Roth, "Fairness in criminal justice risk assessments: The state of the art," Sociological Methods & Research, vol. 50, no. 1, pp. 3-44, 2018.