Diabetes Prediction Analysis.

Name: Rebecca Agbolade
Student Id: D3077427
Course: AI Ethics and Application
Word Count:2025.

Introduction

Diabetes mellitus poses a significant global health challenge, prompting exploration into machine learning for predictive analytics. The integration of machine learning techniques offers promise in refining diabetes prediction and management strategies. However, the pervasive issue of bias within AI (Artificial Intelligence) algorithms demands meticulous consideration, as it can significantly impact human lives.

Numerous studies have highlighted the presence of bias in AI algorithms, particularly in healthcare applications such as diabetes prediction. These biases can stem from various sources, including skewed training data, algorithmic limitations, and societal biases embedded in historical data (Ziad Obermeyer et al., 2019). As a result, the predictive models generated by AI systems may exhibit disparities and inaccuracies that disproportionately affect certain demographic groups.

The consequences of bias in AI-driven diabetes prediction extend beyond algorithmic performance metrics to real-world implications for human health outcomes. For instance, biased algorithms may lead to misdiagnosis, inadequate treatment recommendations, and disparities in access to care for marginalized communities (Caruana et al., 2015). Such outcomes not only jeopardize individual well-being but also exacerbate existing healthcare inequalities and perpetuate systemic injustices.
In this context, our research seeks to address the dual challenges of bias in AI and its impact on human health. By examining the existence of bias in diabetes prediction algorithms and elucidating its effects on human populations, we aim to develop more equitable and reliable predictive models. Through a combination of rigorous analysis, ethical considerations, and innovative methodologies, we endeavour to mitigate bias and improve the effectiveness of AI-driven healthcare interventions.
In summary, this introduction establishes the presence of bias in the AI focus area of diabetes prediction and underscores its profound implications for human health outcomes. By acknowledging these challenges and committing to ethical and equitable research practices, we strive to advance the field towards more inclusive and impactful healthcare solutions.

Data Pre-processing and Model Development

In this project, a comprehensive exploration and pre-processing of the dataset was caried out to ensure data integrity and reduce data redundancy. The aim was to understand the dataset's characteristics, handle any inconsistencies, and transform it into a suitable format for machine learning algorithms.

Dataset Description

The dataset used in this research was obtained from Kaggle and it consists of various attributes related to individuals with diabetes.

	A	B	C	D	E	F	G	H	I
1	gender	age	hypertens	heart_dise	smoking_h	bmi	HbA1c_lev	blood_glu	diabetes
2	Male	18	0	0	never	24.78	8.2	130	1
3	Female	78	0	0	not curren	27.32	5.8	126	1
4	Female	80	0	0	former	27.47	8.8	126	1
5	Female	72	1	0	No Info	27.32	6.5	200	1
6	Male	77	0	0	No Info	27.32	5.8	300	1
7	Female	65	0	0	never	33.41	6.8	126	1
8	Male	80	0	0	No Info	27.32	9	220	1
9	Female	41	0	0	never	47.26	5.7	155	1
10	Male	57	0	0	never	28.34	6.6	130	1
11	Male	79	1	0	not curren	36.72	6.8	260	1
12	Male	39	0	0	ever	27.32	8.8	159	1
13	Male	57	0	0	not curren	33.7	7	300	1
14	Male	55	0	0	No Info	27.32	8.8	200	1
15	Male	64	1	1	current	28.35	6	130	1
16	Female	80	0	0	never	27.32	6	159	1
17	Female	53	0	0	never	50.88	8.2	200	1
18	Female	80	1	0	never	27.32	6.5	220	1
19	Male	54	0	0	never	39.85	6.5	130	1
20	Male	53	0	0	never	33.62	8.8	159	1
21	Female	80	1	1	never	27.32	5.8	280	1
22	Female	37	0	0	never	33.17	7	220	1
23	Male	79	0	0	former	27.32	7.5	126	1
24	Female	69	0	0	never	28.87	5.7	145	1
25	Male	80	0	0	former	26.25	6.1	220	1

Fig 1: Dataset Overview.

The dataset consists of 9 columns and 6774 rows where each attributes includes Gender, Age, Hypertension, Heart disease, Smoking history, BMI, Hba1c_level, Blood glucose, and Diabetes (Diabetes being the Target Variable)

Gender, being a protected variable in this dataset due to its ties to fairness considerations, requires ensuring gender equity in the assessment and modelling of Diabetes predictions to mitigate potential biases resulting from gender discrimination.

Data Exploration

Visualising the distribution of categorical variables such as 'gender' and 'smoking_history', along with the target variable 'diabetes', provided valuable insights into their frequencies and proportions. Histograms were used to visualize the distribution of numerical features, aiding in understanding their spread.

Fig 2: Distribution of categorical variable.

Data Cleaning

Handling Missing Values & Duplicates

While analysing the dataset I realised that the data had no missing values but had 9 duplicate rows which were removed from the dataset. This step ensured that there was no redundancy.

Data Encoding

Categorical variables were converted into numerical format using Label Encoder. This transformation was necessary as machine learning algorithms typically require numerical inputs.

Multi-Collinearity Analysis

An analysis of multi-collinearity was conducted using correlation analysis, revealing no significant correlations among the variables. This indicates a low level of redundancy in the dataset, which improves the effectiveness of predictive modelling. Identifying highly correlated feature pairs above a predefined threshold was important to avoid multicollinearity issues in subsequent modelling. The threshold was set to 0.6 and iterated through the correlation matrix to identify and display highly correlated feature pairs.

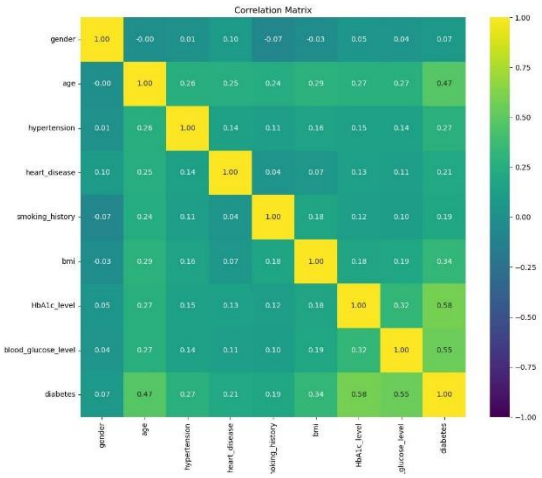


Fig 3: Correlation Analysis.

Data Splitting

Following the correlation analysis, the dataset was divided into training and testing sets at an 80-20 ratio. This division allows the model to be trained on 80% of the data while evaluating its performance on the remaining 20%, ensuring its ability to generalize to new, unseen data instances. This standardised splitting ratio facilitates reliable performance evaluation, thereby enhancing the model's reliability.

Data Splitting

```
# Splitting the data into training and testing sets
X_train, X_test, Y_train, Y_test = train_test_split(X, Y, test_size=0.2, stratify=Y, random_state=2)


# Checking the shape of the resulting sets
print("Shape of X_train:", X_train.shape)
print("Shape of X_test:", X_test.shape)
print("Shape of Y_train:", Y_train.shape)
print("Shape of Y_test:", Y_test.shape)
```

Fig 4: Data Splitting

Normalisation.

After the data splitting, the input features were normalised using MinMaxScaler, which scales the characteristics to a consistent range, typically between 0 and 1. This normalisation process helps minimise biases resulting from features with different scales, thereby stabilising and improving the effectiveness of the model's training process.

Machine Learning Algorithm

I opted to utilise the Random Forest Algorithm in my analysis, recognising its reputation for ensembling decision trees to bolster robustness and accuracy. Considering my focus on predicting medical diagnoses, specifically Diabetes, I found Random Forest's capacity to combat overfitting and enhance generalisation particularly advantageous for my research. This algorithm aligns well with the complexities inherent in medical data, making it a fitting choice for my specific use case.

Performance Evaluation

Various evaluation criteria and methodologies are employed to assess the performance of the Random Forest model in predicting diabetes. Accuracy provides an overall measure of prediction correctness. The confusion matrix offers a detailed breakdown of predictions, delineating true positives, true negatives, false positives, and false negatives. Additionally, the classification report furnishes a comprehensive evaluation of the model's performance across multiple categories, offering crucial classification metrics such as

precision and recall for each class. The ROC-AUC curve assesses the model's discriminatory capability. These metrics collectively provide valuable insights into the model's performance.

Classification Report

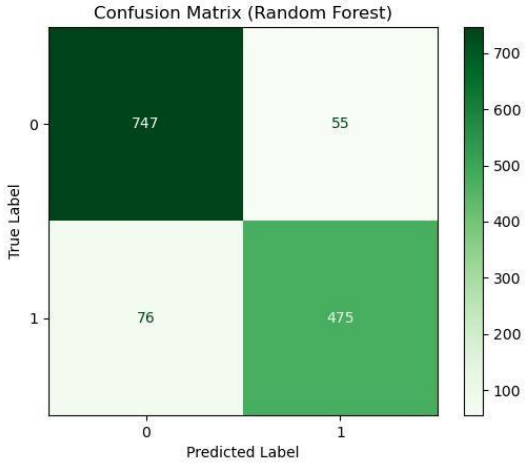
Random Forest Classification Report:

	precision	recall	f1-score	support
0	0.92	0.93	0.92	802
1	0.90	0.88	0.89	551
accuracy			0.91	1353
macro avg	0.91	0.90	0.90	1353
weighted avg	0.91	0.91	0.91	1353

Fig 5: Classification Report.

After training the model, the trained model was used on the testing data to evaluate how well the model can predict the unknown. After evaluating the model on the test data, the classification report was printed out. The classification report shows an accuracy of 91%.

Confusion-Matrix



The confusion matrix provides crucial insights into the model's performance by revealing the frequency of classifications made by the algorithm. Specifically, the model accurately identified 747 instances as negative (True Negative), indicating its proficiency in detecting negative outcomes. However, there were 55 instances where negative outcomes were incorrectly identified as positive (False Positive). Moreover, 76 positive outcomes were mistakenly classified as negative (False Negative), while 475 cases were correctly classified as positive (True Positive), displaying the model's precision in recognizing claims. Analyzing these figures within the confusion matrix yields valuable insights into the effectiveness of the model.

Receiver Operating Characteristic (ROC) Curve

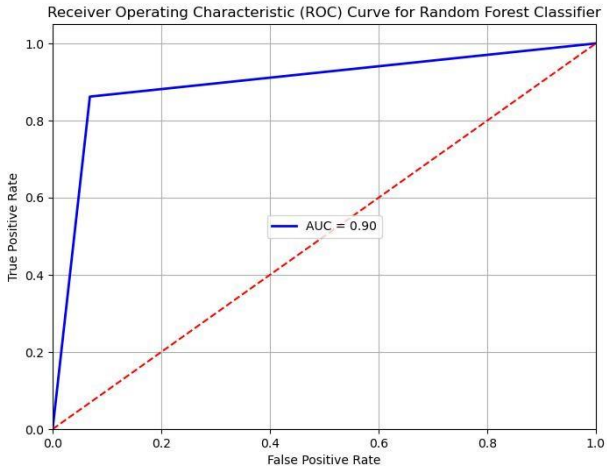


FIG 6: ROC Curve.

An AUC of 0.90 indicates strong discriminating ability for the model, meaning it effectively distinguishes between positive and negative outcomes with high accuracy. This high AUC value reflects the model's reliability and robust performance in classification tasks.

Bias & Fairness

The model's performance was assessed using a range of evaluation metrics, encompassing accuracy, precision, recall (equal opportunity), and demographic parity.
Special attention was paid to fairness and equity considerations by comparing model

performance across gender groups, with a focus on men and women.

Model Performance on Males

Evaluating the performance of the random forest model on the male group. This evaluation shows an accuracy of 91%, with a recall of 89% and precision of 93%.

```
Calculated Accuracy =  0.9186440677966101
Calculated Recall =  0.8917910447761194
Calculated Precision =  0.9263565891472868
```

The classification report shows that the model's performance in the male group had a positive rate of 44 %.

```
Random Forest Classification Report:
              precision    recall  f1-score   support

           0       0.91      0.94      0.93       322
           1       0.93      0.89      0.91       268

    accuracy                           0.92       590
   macro avg       0.92      0.92      0.92       590
weighted avg       0.92      0.92      0.92       590

Positive Rate 0.43728813559322033
```

Fig 7: Classification Report for Male Group

The confusion matrix for the male group shows the following performance: True Positive as 238, True Negative as 301, False Positive as 21 and
False Negative as 30.

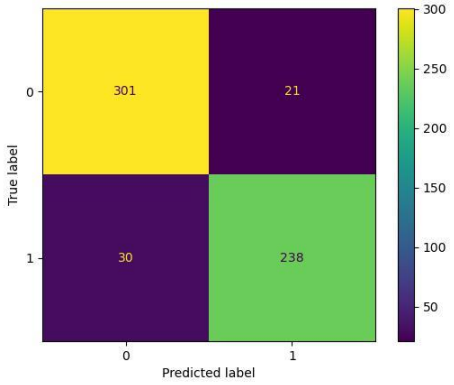


Fig 8: confusion Matrix for Male Group.

Model Performance on Females

Evaluating the performance of the random forest model on the female group. This evaluation shows an accuracy of 91%, with a recall and precisionof88% and Precision of 89%

```
Calculated Accuracy =  0.9121887287024901
Calculated Recall =  0.8763250883392226
Calculated Precision =  0.8857142857142857
```

The classification report shows that the model's performance in the female group had a positive rate of 37%.

```
Random Forest Classification Report:
              precision    recall  f1-score   support

           0       0.93      0.93      0.93       480
           1       0.89      0.88      0.88       283

    accuracy                           0.91       763
   macro avg       0.91      0.90      0.91       763
weighted avg       0.91      0.91      0.91       763

Positive Rate 0.3669724770642202
```

Fig 9: Classification Report for Female Groups.

The confusion matrix for the male group shows the following performance: True Positive as 244, True Negative as 443, False Positive as 37 and
False Negative as 39.

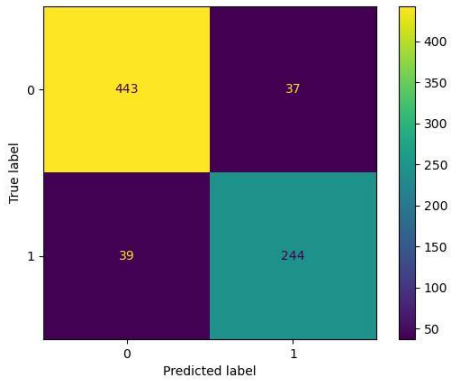


Fig 10: confusion Matrix for Female Group**.**

Performance Evaluation Metrics

1. Equal Accuracy:

Equal accuracy measures whether the model performs equally well across different demographic groups, irrespective of sensitive attributes such as gender.

$$\frac{TP + TN}{TP + TN + FP + FN}$$

Evidence:

Equal Accuracy for Men: 0.908
Equal Accuracy for Women: 0.907
Equal Opportunity (Recall) across the protected group.

```
# Assigning values for males
TN_m = 297
FP_m = 25
FN_m = 29
TP_m = 239

# Assigned values for females
TN_f = 446
FP_f = 34
FN_f = 37
TP_f = 246

# Calculate equal accuracy for males
total_predictions_m = TN_m + FP_m + FN_m + TP_m
equal_accuracy_m = (TP_m + TN_m) / total_predictions_m

# Calculate equal accuracy for females
total_predictions_f = TN_f + FP_f + FN_f + TP_f
equal_accuracy_f = (TP_f + TN_f) / total_predictions_f

# Output the results
print("Equal Accuracy for Men:", equal_accuracy_m)
print("Equal Accuracy for Women:", equal_accuracy_f)
```

Analysis:

The difference in equal accuracy between men and women is minimal, indicating similar performance of the model across gender groups. This suggests that, overall, the model achieves comparable accuracy in predicting diabetes for both men and women, providing initial evidence of fairness in predictive performance.

Existence of Bias:

The minimal difference in equal accuracy between men and women (Men: 90.85%, Women: 90.69%) suggests relatively equitable predictive performance across gender groups. However, the presence of bias cannot be definitively ruled out based solely on equal accuracy. Other fairness metrics should be considered to provide a more comprehensive assessment of bias.

2. Demographic Parity:

Demographic parity assesses whether the proportion of positive outcomes (e.g., true positives) relative to the total number of positive predictions is consistent across different demographic groups.

$$\frac{TP + FP}{TP + TN + FP + FN}$$

Evidence:

Demographic Parity for Men: 0.447
Demographic Parity for Women: 0.366

```
# Assigning values for males
TN_m = 297
FP_m = 25
FN_m = 29
TP_m = 239

# Assigned values for females
TN_f = 446
FP_f = 34
FN_f = 37
TP_f = 246

# Calculating Demographic Parity for Males
total_predictions_m = TN_m + FP_m + FN_m + TP_m
Demograpghic_Parity_m = (TP_m + FP_m)/ total_predictions_m

# Calculating Demographic Parity for Females
total_predictions_f = TN_f + FP_f + FN_f + TP_f
Demograpghic_Parity_f =(TP_f + FP_f)/ total_predictions_f


# Printing the result
print("Demographic Parity for Males;", Demograpghic_Parity_m)
print("Demographic Parity for Females;", Demograpghic_Parity_f)
```

```
Demographic Parity for Males; 0.44745762711864406
Demographic Parity for Females; 0.3669724770642202
```

Analysis:

There is a slight disparity in demographic parity between men and women.
The proportion of true positives among positive predictions is slightly higher for men compared to women, suggesting potential differences in the model's treatment of gender groups.
This indicates a potential bias in the model's predictions, where men may receive slightly more favourable outcomes than women.

Existence of Bias:

The slight disparity in demographic parity between men and women (Men: 44.7%, Women: 36.6%), where men exhibit a slightly higher proportion of true positives among positive predictions, indicates potential bias in the model's predictions. This suggests that the model may favour one gender group over the other in terms of positive outcomes, which could lead to inequitable treatment of gender groups.

3. Recall (Equal Opportunity):

Recall (Equal Opportunity) measures the proportion of actual positive cases (e.g., true positives) correctly identified by the model for each demographic group.

$$\frac{TP}{TP + FN}$$

Evidence:

Recall (Equal Opportunity) for Men: 0.892
Recall (Equal Opportunity) for Women: 0.869

```
]:
# Recall for Men (True Positive Rate)
recall_m = TP_m / (TP_m + FN_m)

# Recall for Women (True Positive Rate)
recall_f = TP_f / (TP_f + FN_f)

# Calculate recall (equal opportunity) for men and women
recall_positive_m = recall_m
recall_positive_f = recall_f
print("Recall (Equal Opportunity) for Men:", recall_positive_m)
print("Recall (Equal Opportunity) for Women:", recall_positive_f)
```

```
Recall (Equal Opportunity) for Men: 0.8917910447761194
Recall (Equal Opportunity) for Women: 0.8692579505300353
```

Analysis:

The recall for men is slightly higher than for women, indicating a potential discrepancy in correctly identifying positive cases across gender groups.
This suggests that the model is slightly better at correctly identifying positive cases for men compared to women, when considering equal treatment or opportunity.
The difference in recall rates may imply bias in the model's predictions, favouring one gender group over the other in terms of correctly identifying diabetes cases.

Existence of Bias:

The difference in recall rates between men and women (Men: 89.28%, Women: 86.93%), with men having a slightly higher recall than women, hints at potential bias in the model's predictions. A higher recall rate for one gender group suggests that the model may be better at correctly identifying positive cases for that group, potentially resulting in unequal treatment, or missed opportunities for the other gender group.

Conclusion

While the model demonstrates comparable accuracy across gender groups, disparities in demographic parity and recall rates suggest the presence of bias in the model's predictions. The slight differences observed in fairness metrics indicate potential inequities in the model's treatment of gender groups. Further investigation and mitigation strategies, such as balanced sampling, fairness-aware algorithms, and model calibration, are warranted to address and mitigate bias in the model's predictions, ensuring fairness and equity across demographic groups.

References

1. Rajkomar, A., Oren, E., Chen, K., Dai, A. M., Hajaj, N., Hardt, M., ... & Zhang, M. (2018). Scalable and accurate deep

learning with electronic health records. npj Digital Medicine, 1(1), 1-10.
2. Beam, A. L., & Kohane, I. S. (2018). Big data and machine learning in health care. JAMA, 319(13), 1317-1318.
3. Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. Science, 366(6464), 447-453.
4. Caruana, R., Lou, Y., Gehrke, J., Koch, P., Sturm, M., & Elhadad, N. (2015). Intelligible models for healthcare: Predicting pneumonia risk and hospital 30-day readmission. Proceedings of the 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 1721-1730.