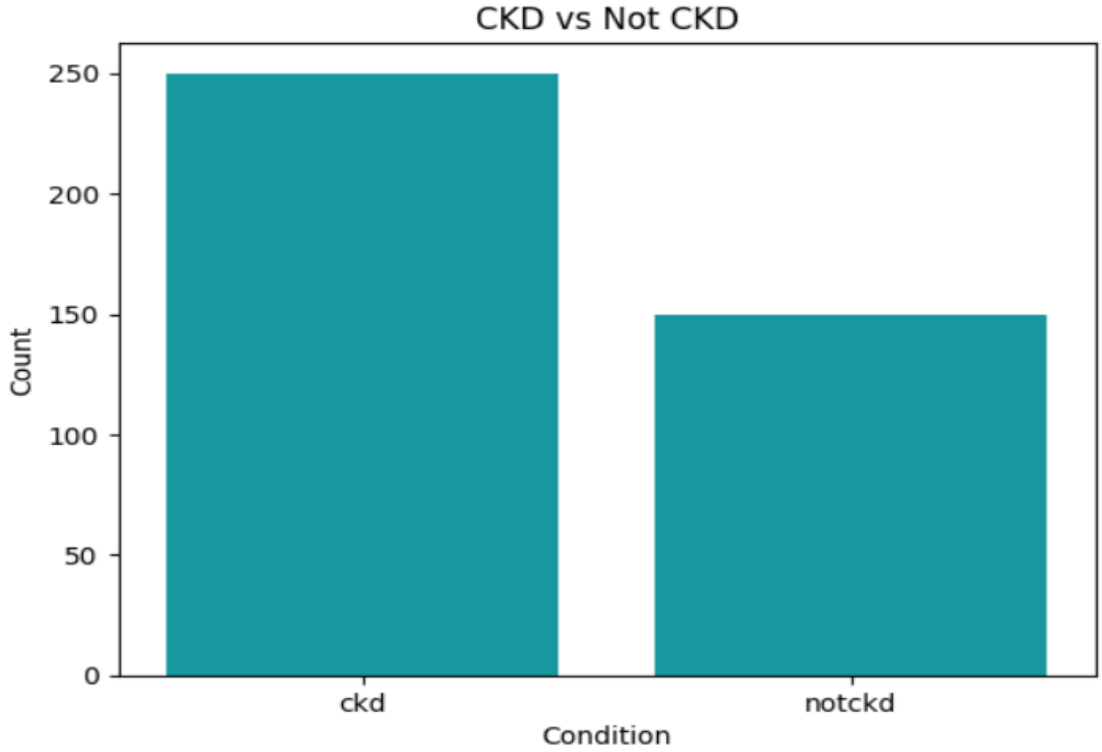
Chronic Kidney Disease – Visual Analysis

Exploratory Data Analysis & Model Results

Section 1 – Data Distribution & Balance

Figure 1: CKD vs Non-CKD Bar Chart

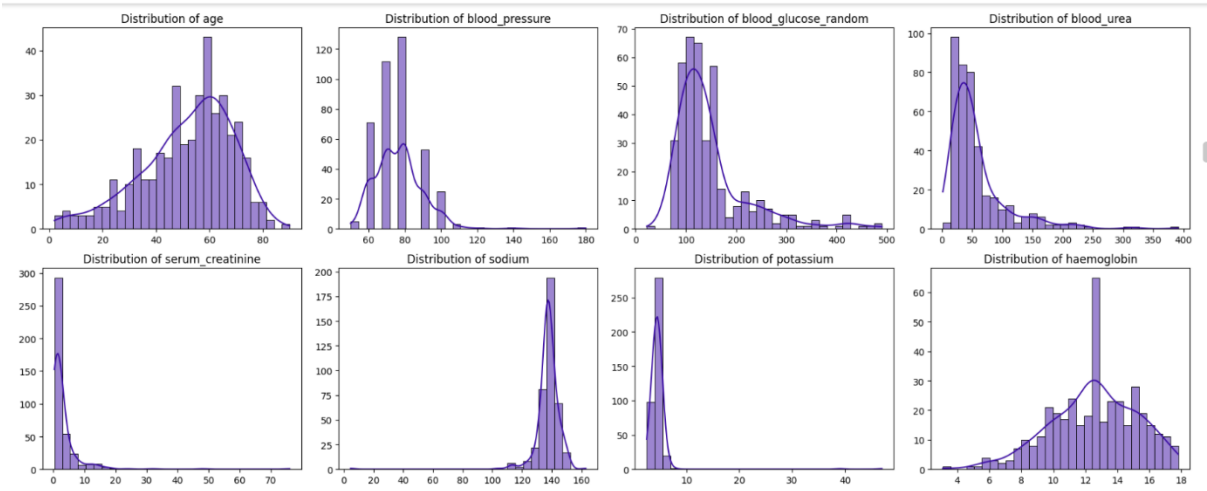
Shows the distribution of patients with CKD compared to those without. Reveals class imbalance, which must be considered during model training.



Section 2 – Feature Distributions

Figure 2: Histograms of Numerical Features

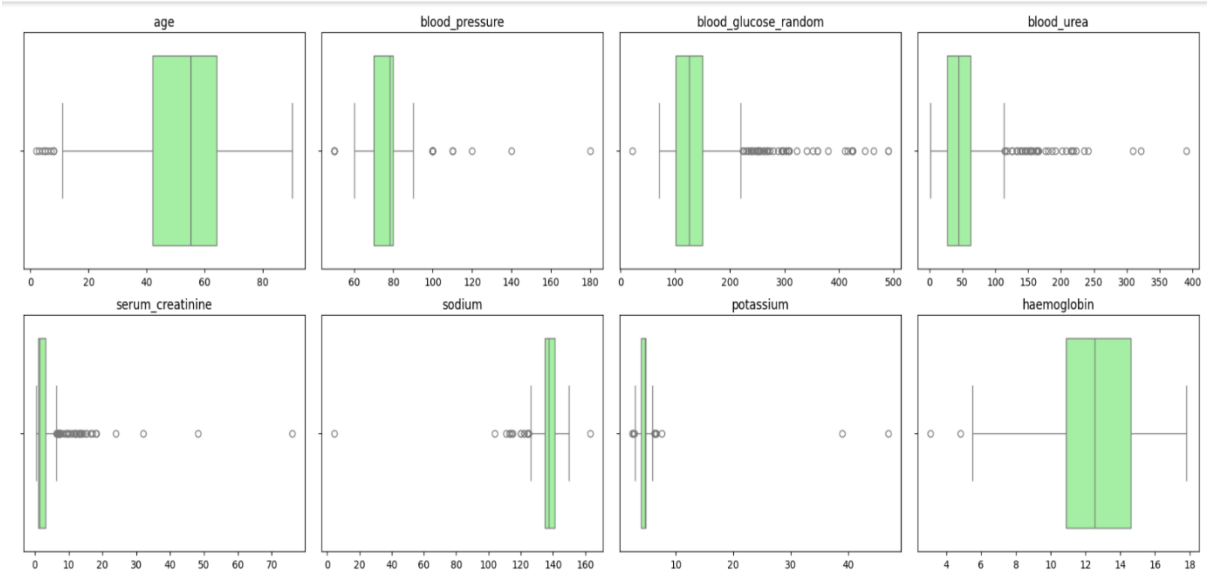
Visualizes spread and skewness of lab values like serum creatinine, haemoglobin, and blood urea. Identifies critical markers for CKD detection.



Section 3 – Outlier Detection

Figure 3: Boxplots of Key Numerical Features

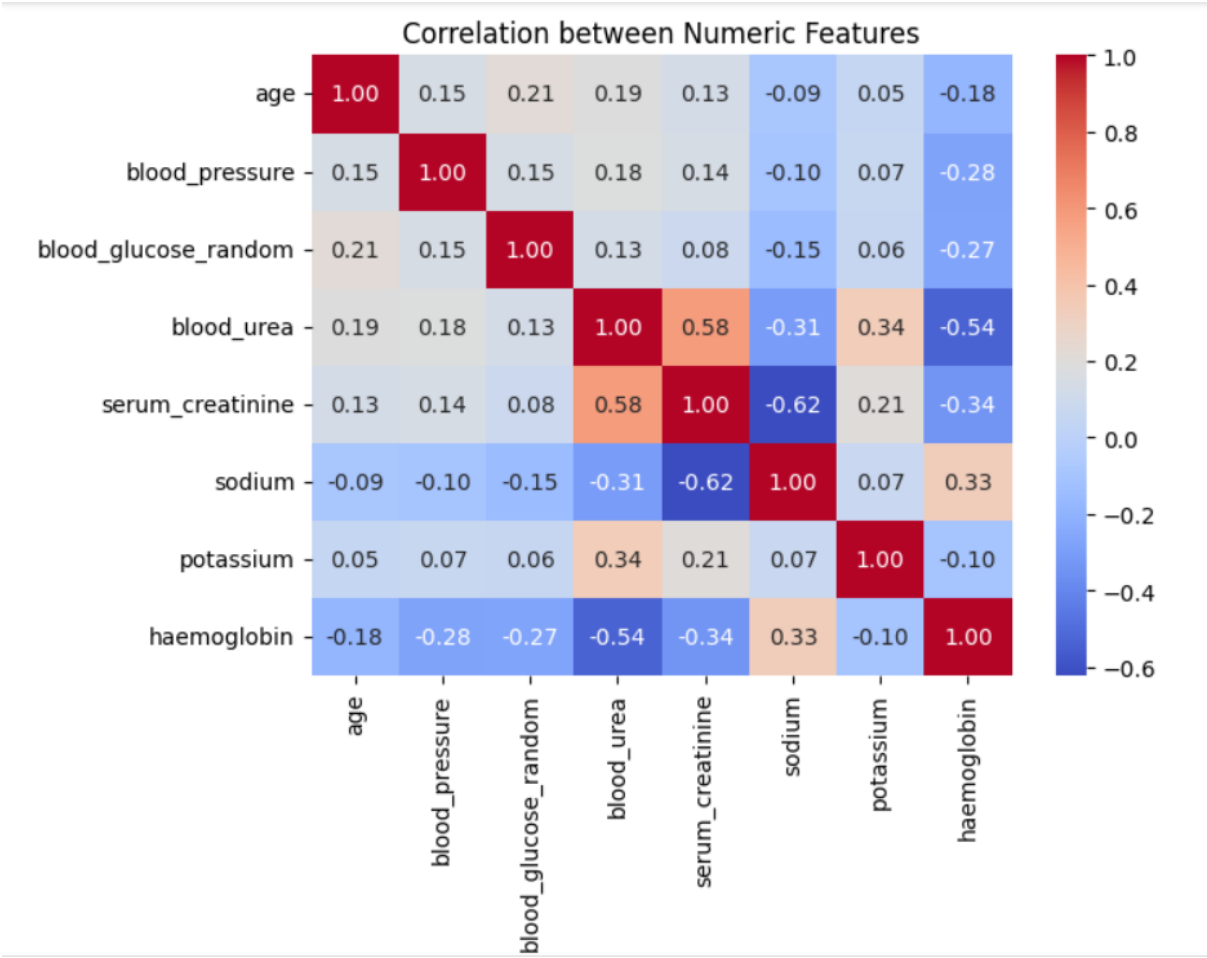
Highlights presence of extreme values in features such as blood urea and serum creatinine. Outlier handling was required to improve model reliability.



Section 4 – Correlations

Figure 4: Correlation Heatmap

Strong positive correlation of serum creatinine & blood urea with CKD. Haemoglobin shows a negative correlation with CKD cases.



Section 5 – Bivariate/Multivariate Insights

Figure 5: Violin Plot of Numerical Features

Lower haemoglobin levels strongly associated with CKD patients.

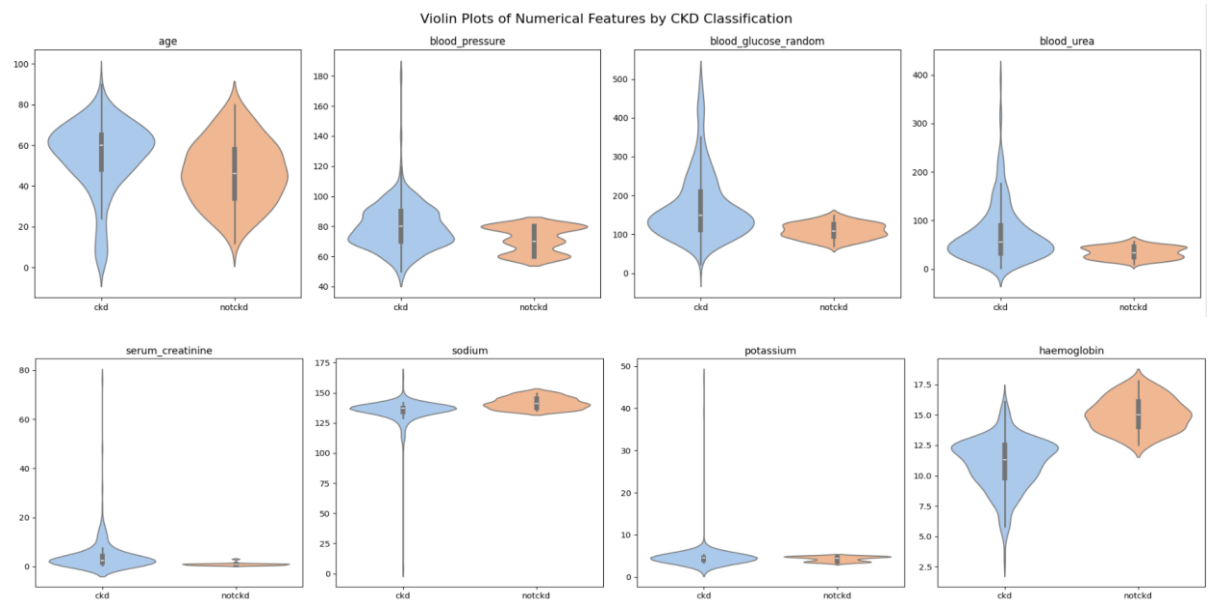


Figure 6: Barplots for categorical features

Shows how comorbidities like hypertension and diabetes increase likelihood of CKD.

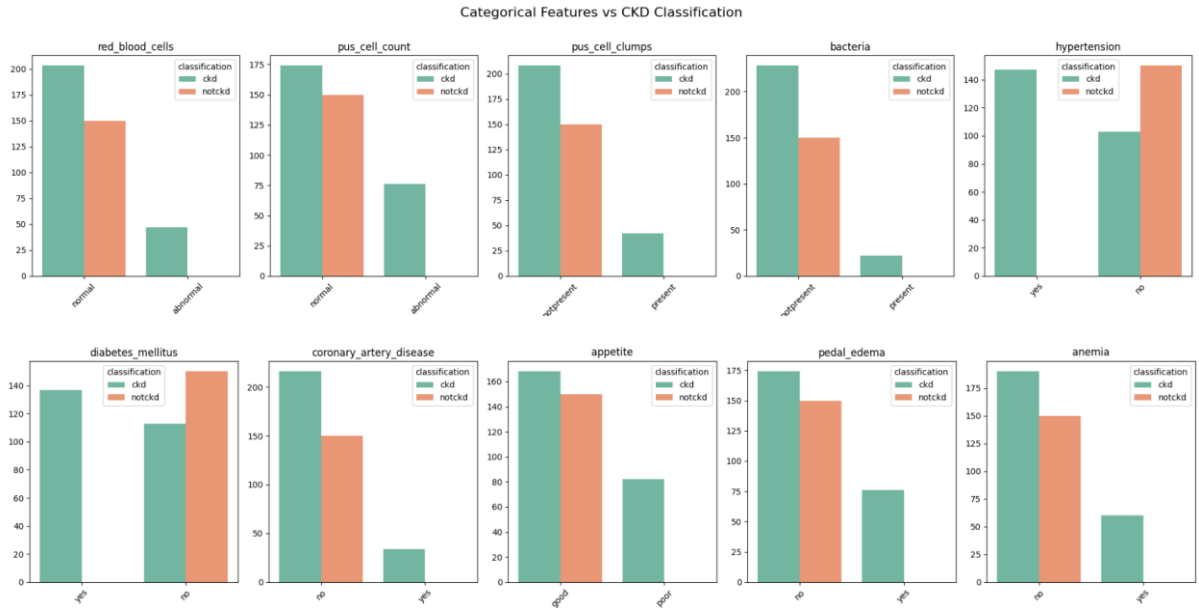
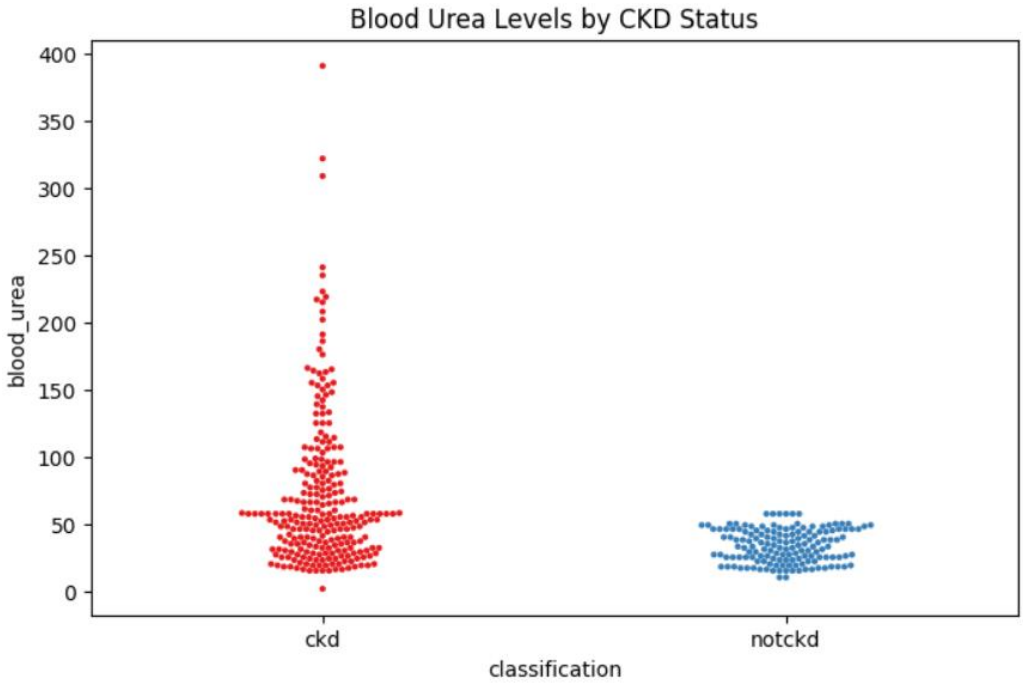


Figure 7: Swarmplots

Swarmplots helped visualize point-wise variations in blood urea and serum creatinine by CKD classification.



Section 6 – Model Results

Figure 7: Accuracy Comparison of Different Models

SGBoost selected as final model with accuracy 98.7%, balancing generalization and performance.

Section 7 – Summary

The visual analysis confirmed that lab values (serum creatinine, albumin, haemoglobin) and comorbidities (hypertension, diabetes) are key indicators of CKD. These insights guided model feature selection and improved interpretability for medical relevance.