STROKE PREDICTION

AI2 - PROJECT - 2024-2025 - CS

STROKE
PREDICTION

STROKE PREDICTION

INTRODUCTION

This project is a Machine Learning-based Stroke Prediction Application that allows users to input specific health-related parameters to predict the likelihood of a stroke.

STROKE PREDICTION

OUR AIM

Support Early Diagnosis:

Enable healthcare professionals to identify high-risk individuals, allowing timely interventions to prevent strokes.

Promote Healthcare Accessibility:

Offer scalable and efficient solutions for stroke prediction that can be integrated into various medical and insurance systems.

STROKE
PREDICTION

GOAL

Develop a predictive model that accurately identifies the likelihood of stroke occurrence based on demographic and health data

Assist in early detection and prevention of strokes, which is critical for timely medical intervention.

STROKE PREDICTION

CHALLENGES

1. Unpredictable medical costs affect financial stability
2. Need for a predictive model based on demographic and health data.
3. Challenges in accurately pricing insurance policies

STROKE PREDICTION

Methodology Overview

Data Collection &Preprocessing

Dataset overview (healthcare-dataset-stroke-data.csv).

Handling missing data and encoding categorical features

Exploratory Data Analysis

Age, BMI, smoker, and region distributions (visualizations).

Model Training

Linear Regression

Ridge Regression

Lasso Regression

Decision Tree Regression

Random Forest Regression.

Evaluation Metrics

R-squared values for training and test data.

STROKE PREDICTION

ACCURACY

FOR EACH ALGORITHM

Algorithm	Accuracy
Logistic Regression	82.51%
KNN	91.9%
XGBoost	94.38%
Random Forest	95.1%
SVM	88.09%
Decision Tree	91.18%

STROKE PREDICTION

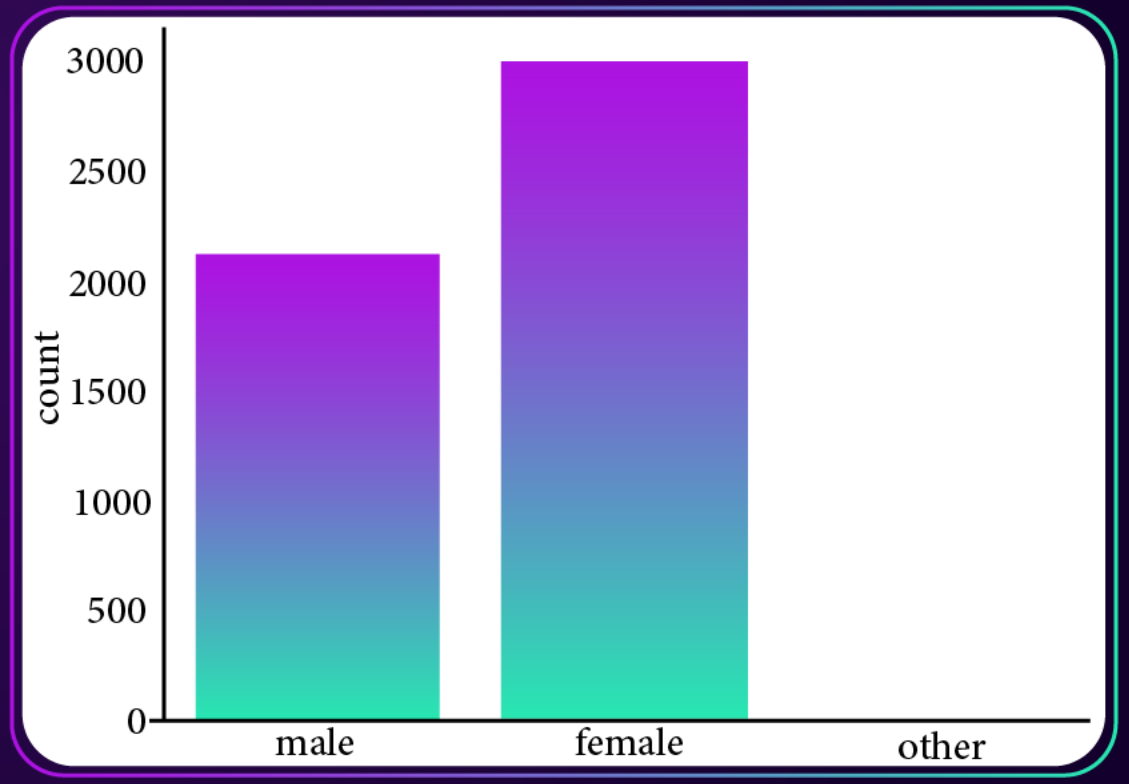
DATA SET

1	id	gender	age	hypertension	heart_disease	ever_married	work_type	Residence_type	avg_glucose_level	bmi	smoking_status	stroke
2	9046	Male	67	0	1	Yes	Private	Urban	228.69	36.6	formerly smoked	1
3	51676	Female	61	0	0	Yes	Self-employed	Rural	202.21	N/A	never smoked	1
4	31112	Male	80	0	1	Yes	Private	Rural	105.92	32.5	never smoked	1
5	60182	Female	49	0	0	Yes	Private	Urban	171.23	34.4	smokes	1

STROKE PREDICTION

PROJECT DATA

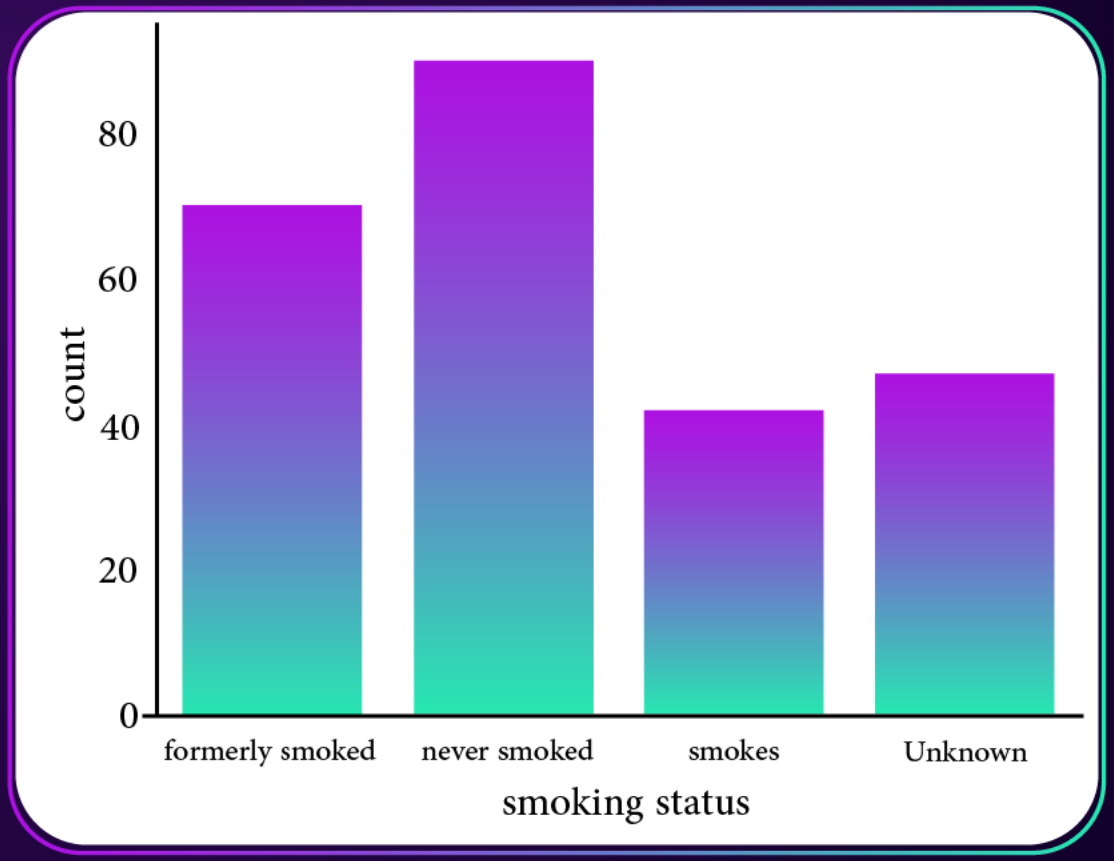
We then visualized the gender distribution using sns.countplot(). This plot shows the number of males and females in the dataset. It's essential to identify whether the dataset is balanced in termsof gender or if there's a disproportionate representation. If the distribution is imbalanced,different approaches may be needed to ensure robust analysis.



STROKE PREDICTION

PROJECT DATA

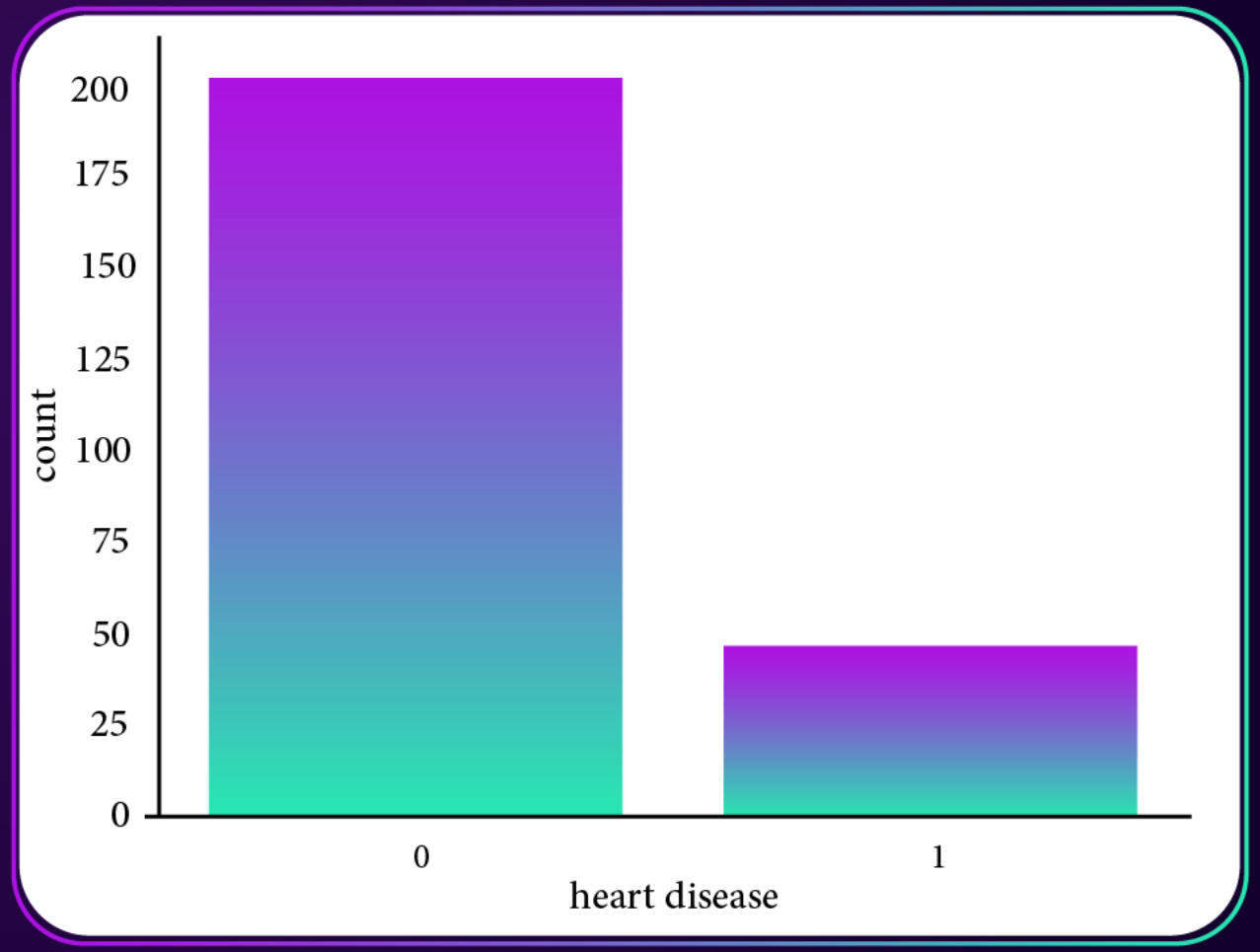
We also explored distributions for Smoking Status: To understand the prevalence of smokers and their potential influence



STROKE PREDICTION

PROJECT DATA

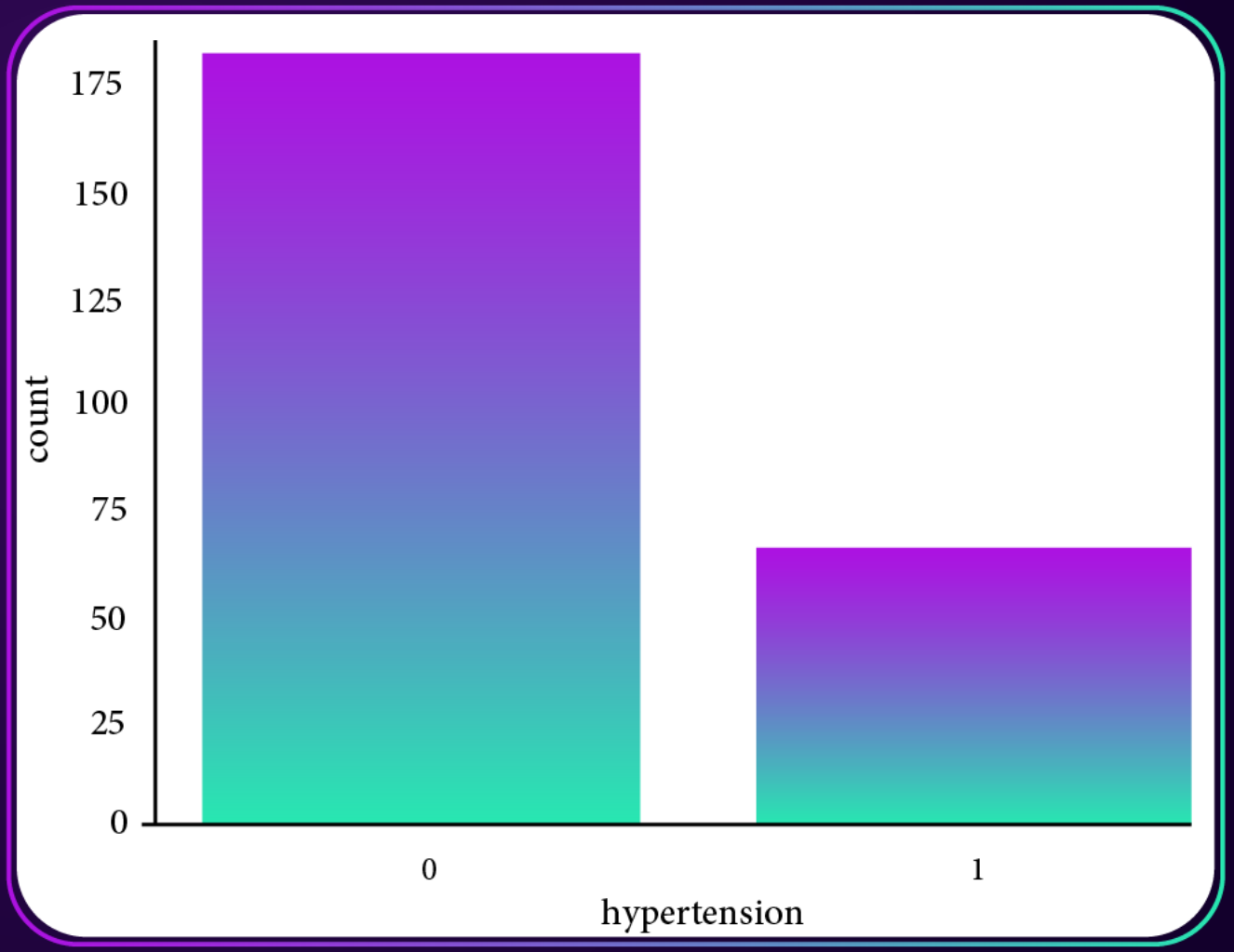
We also explored distributions for heart disease Status



STROKE PREDICTION

PROJECT DATA

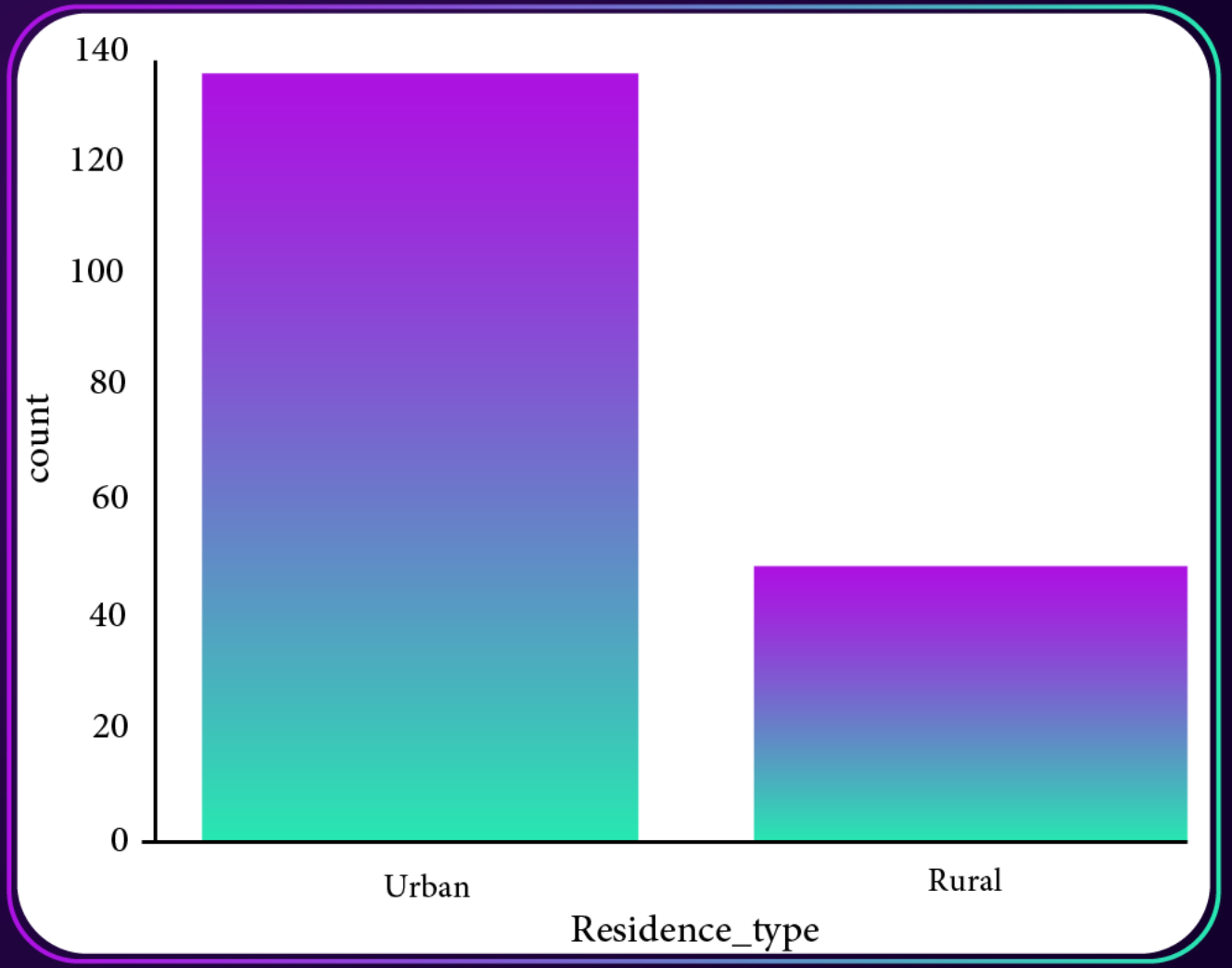
We also explored distributions for hypertension Status



STROKE PREDICTION

PROJECT DATA

We also explored distributions for Residence type Status



STROKE PREDICTION

TEAM

AHMED MOHAMED GALAL

AMR AHMED ZAKARIA

GAMAL MOHAMED ALI

EMAN MOHAMED KHATER

ESRAA MORSI MORSI