

# Stroke Prediction

Priscilla E. Cole

# Who is my our Stakeholder? What is our Problem?

Our stakeholder holds

According to the World Health Organization, stroke is the 2nd leading cause of death. This dataset is used to predict the likelihood of a patient to have a stroke based on other parameters.

Our stakeholder has challenged us to use these parameters to select a model that would best aid in the prediction of the likelihood of a patient suffering a stroke and how to implement a model accordingly.

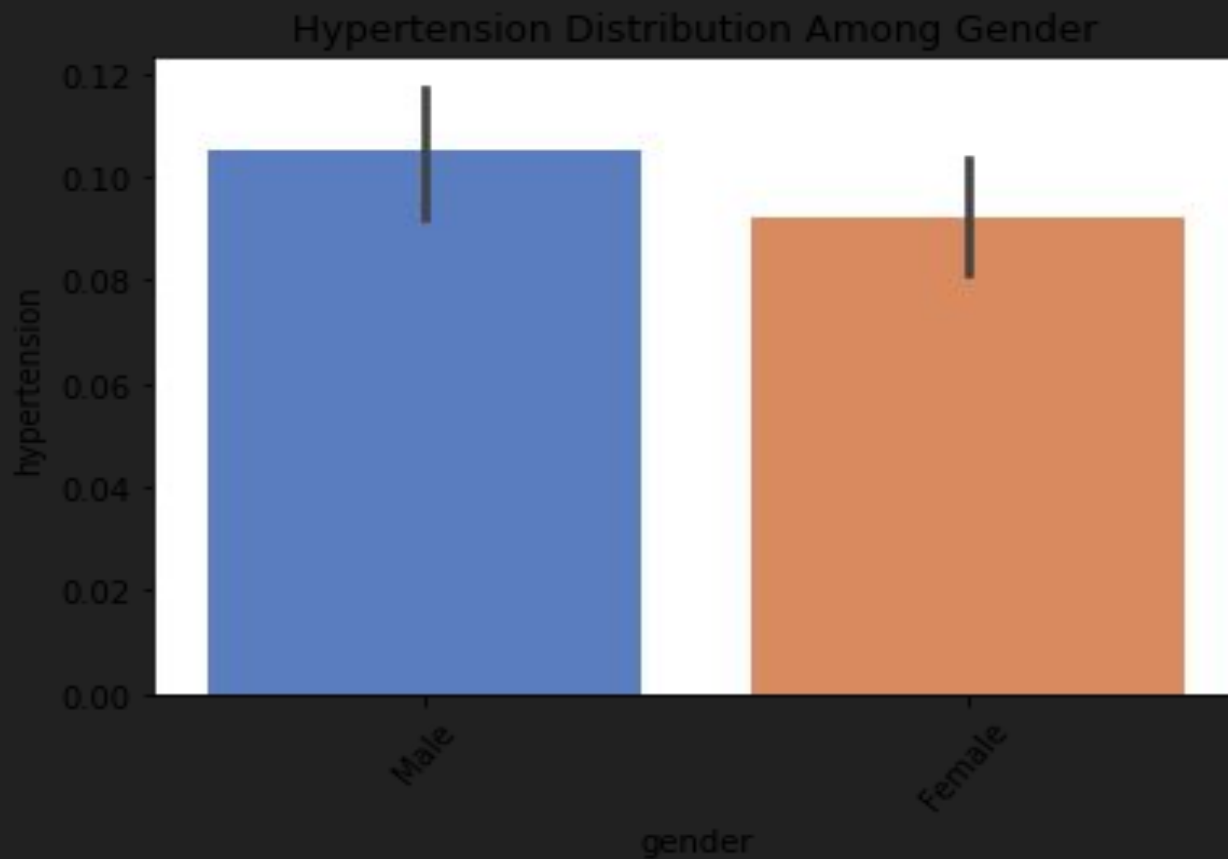
# Introduction to the Data

This classification dataset is provided by Kaggle.

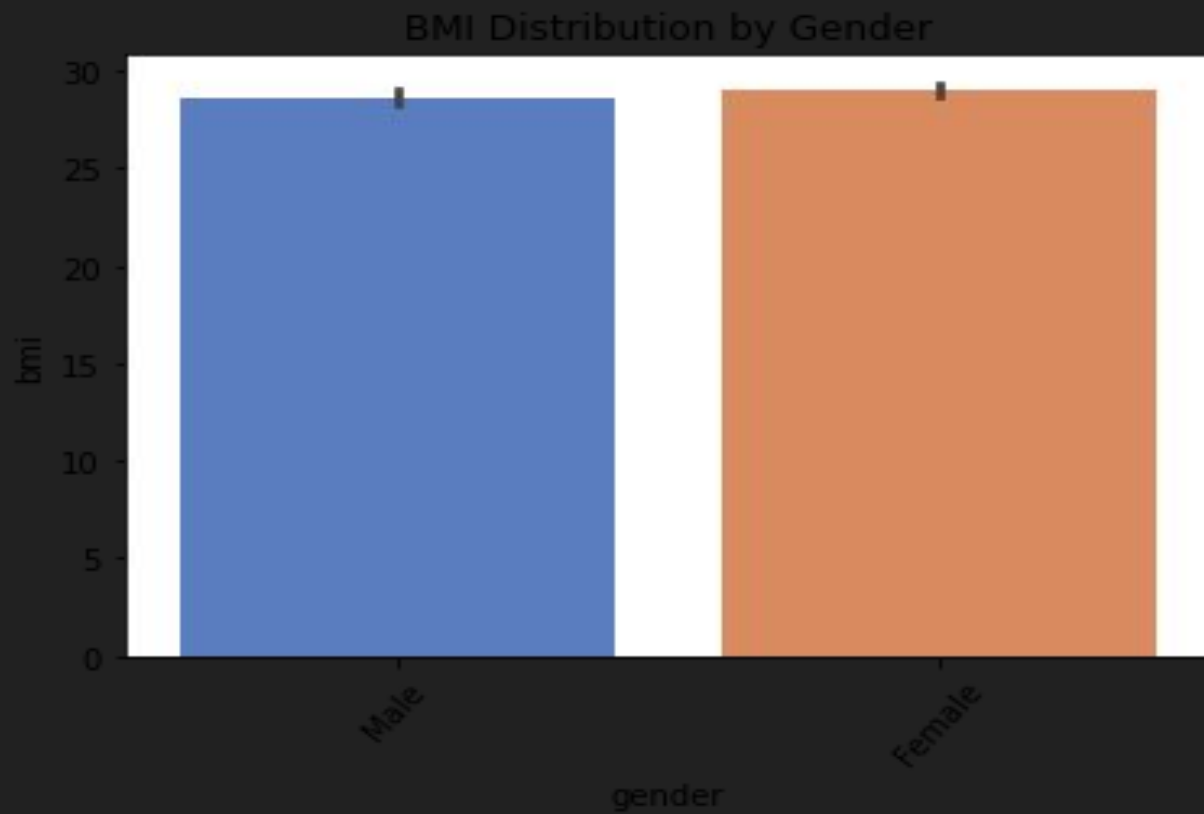
Link: <https://www.kaggle.com/datasets/fedesoriano/stroke-prediction-dataset>

This dataset is comprised of 5,110 observations and 12 attributes.

# Hypertension and Gender



# BMI and Gender



# Strengths and Limitations

A strength of our model was that our model confirmed what the class balance originally showed. Prior to any modeling, the class balance showed approximately 95% chance of no stroke and a 5% chance of stroke based on the attributes given. When evaluating the accuracy of all of the models, those predictions were right near that same baseline, even after hypertuning.

A limitation is that while our modeling did show us that not many changes need to be made, there were still some showings for false negatives, which we really do not want.

# False Positives and False Negatives

A false positive would mean that the model predicted that a patient would have a stroke and that person did not have a stroke. While it is not ideal to misdiagnose someone, a false positive means that the patient would be prescribed the medication, given the treatment and health information to recover from and prevent further strokes in the future.

A false negative indicates that it was predicted that a patient would not have a stroke when in actuality, that patient did have a stroke. A false negative would be worse for our stakeholders and business because not correctly predicting and/or diagnosing a stroke would mean that a patient does not get the proper treatment, medications or information to correct and improve their health.

# Final Recommendations

After preparing and modeling the data, we elected to move forward with implementing the KNN model classifier with PCA. While it met near the balance baseline (shown from the class balance), the average macro f1 evaluation also scored very well when compared to the models.

While there may not be big, overhaul changes that need to be made, there are some things that need to be fine tuned and addressed (like having the least amount of false negatives) and our chosen model will definitely get us to that level.