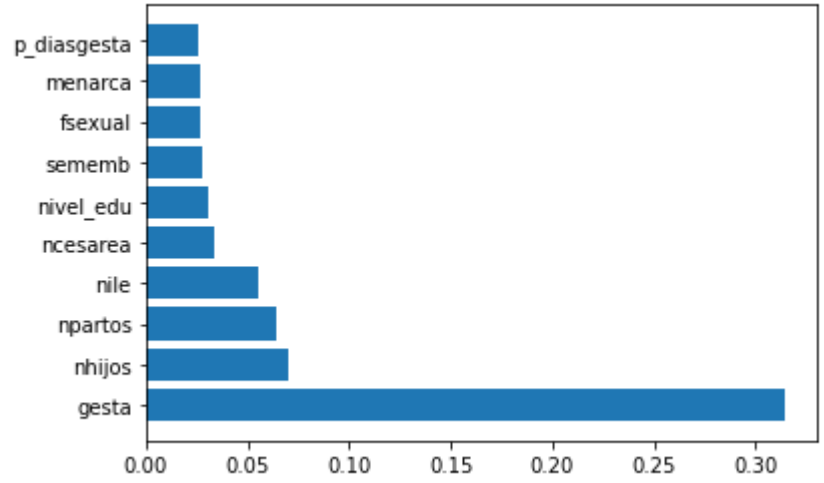
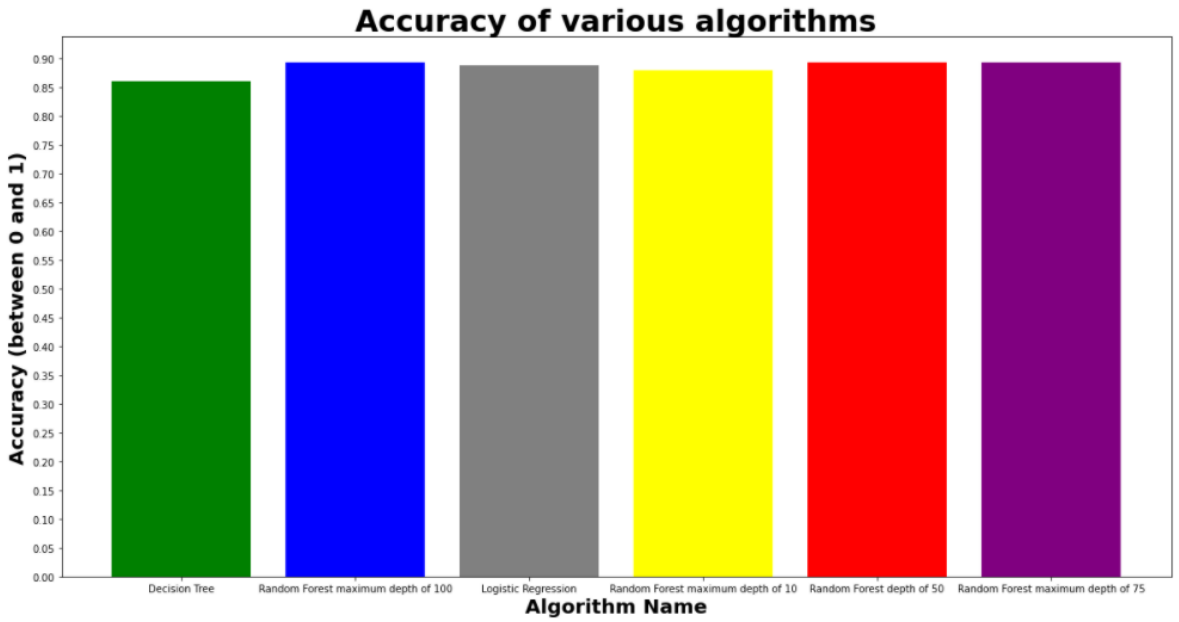
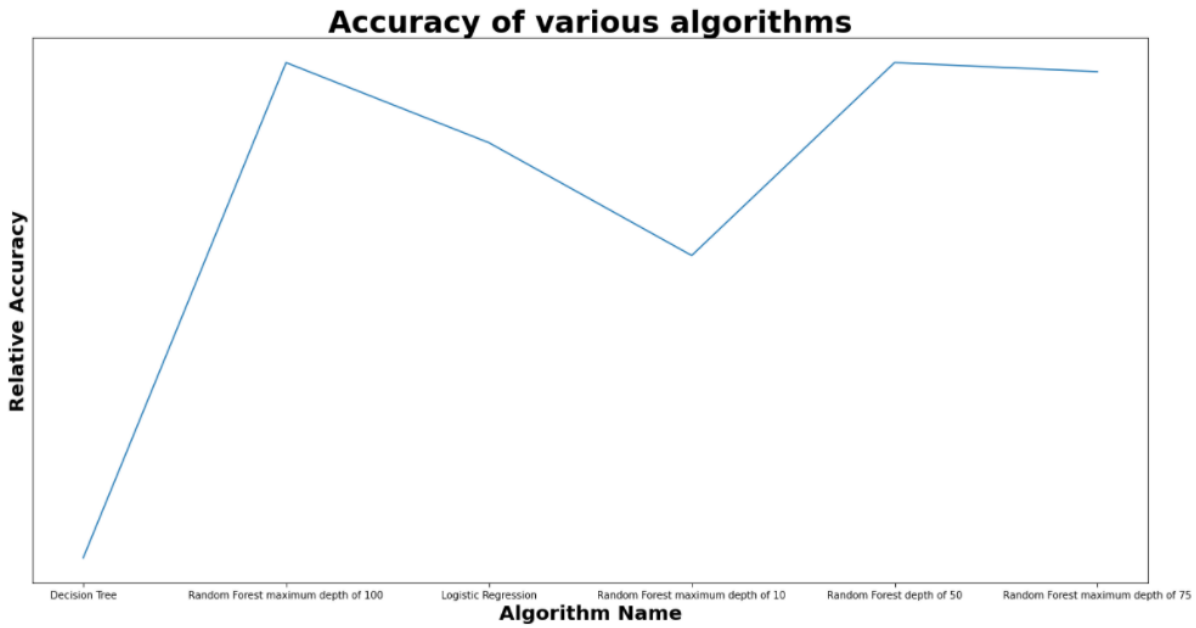
ABORTIONS WITH MACHINE LEARNING PROJECT

I first split the ILE dataset into training data and testing data. I then created the models to use in order to predict the number of abortions - a decision tree, random forests, and a logistic regression model and trained them using the training dataset. After training, each model predicted the number of abortions on the testing dataset. Then the accuracy of each model was determined by comparing the predicted number of abortions for each row in the testing dataset to the actual number of abortions for each row in the testing dataset.



The five most beneficial factors in predicting the number of abortions (in order) are gesta, the number of children they have, the number of illegal abortions, npartos and the number of c-sections.





The random forest with a maximum depth (maximum distance between the root node and a leaf node) of 100 and 50 had the highest accuracy out of all the models at 89.3%, whereas the decision tree had the lowest accuracy out of all the models at 85.99%.The random forest with a maximum depth of 75 had the next highest accuracy (at 89.24%), followed by the logistic regression model (at 88.76%), followed by the random forest with a maximum depth of 10 (at 88%).