**Machine Learning Assignment – 2 Report**

**Uday Bhaskar Valapadasu**

**11696364**

The study used logistic regression to predict diabetes on a dataset with 768 samples and 9 features. Three data splits were tested: 80-20, 70-30, and 60-40 (train-test).

**Key Findings:**

1. The 80-20 and 60-40 splits both achieved the highest accuracy of 75.32%.
2. The 80-20 split was selected for further analysis.
3. For the 80-20 split:
   1. Accuracy: 75.32%
   2. AUC: 0.8165
   3. Confusion Matrix: [[80 19], [19 36]]

**Bootstrap Analysis (80-20 split):**

* 1000 iterations performed
* P-value: 0.51
* 95% Confidence Interval: [0.708, 0.792]

The model shows good predictive performance with an AUC of 0.8165, indicating strong discrimination ability. The bootstrap analysis suggests the accuracy estimate is robust, with a narrow confidence interval.

A black screen with white text

Description automatically generated

A black screen with white text

Description automatically generated

A black screen with white text

Description automatically generated

