



# CALIFORNIA SCIENCE & ENGINEERING FAIR 2018 PROJECT SUMMARY

<b>Name(s)</b> Tejas N. Rao	<b>Project Number</b> <b>S0827</b>
<b>Project Title</b> <b>Logistic Regression and Decision Tree ML Algorithms to Predict Type-2 Diabetes</b>	
<p style="text-align: center;"><b>Abstract</b></p> <p><b>Objectives/Goals</b> Compare two Statistical Models to predict Type-2 Diabetes - Logistic Regression and Decision Trees. Determine which patient attributes - Age, Body Mass Index, Glucose Concentration, Genetics, % of time pregnant are most significant for Diabetes</p> <p>Determine the following for each model to aid comparison: Accuracy, Sensitivity, Specivity, ROC Area under Curve.</p> <p>Build a simple web application to use the model in mobile phones. Application should accept key patient data and return probability of diabetes Application should run on phone and browser.</p> <p><b>Methods/Materials</b> UC Irvine Department of Machine Learning Pima Indians Diabetes DataSet. This dataset provides details on 782 Pima Indians for Age, BMI, Pregnancy etc. Scikit-learn: Machine learning in Python Logistic Regression and Decision Tree algorithm packages in Python. Pythonanywhere for Hosting and running Python Applications. Jupyter notebooks running on Azure Cloud.</p> <p><b>Methods</b> Scikit-learn Machine Learning toolkit in Python was used for running Classification Models DataSet has 768 patient records which were divided into 75% (576 records) for Training data and remaining 25% (192 records) for Test data. Both models Logistic Regression and Decision Trees, are Trained and Scored with training data and test data respectively</p> <p>Prediction Accuracy is measured as <math>(TP+TN) / (TP+TN+FP+FN)</math> Sensitivity is measured as <math>TP / (TP+FN)</math> Specificity is measured as <math>TN / (TN+FP)</math> HTML5 was used to build a simple webapp that accepts Patient Data in a Form and calls backend Python App.</p> <p><b>Results</b> Logistic Regression Model has</p>	
<b>Summary Statement</b> Prevent Diabetes using Machine Learning Algorithms- Logistic Regression and Decision Trees	
<b>Help Received</b> Mr Wilke (San Mateo High School), Ms Bharathi Udupi (Oracle)	