**Project Report of the Data Analytics Internship**

(Outlining Work Done and Results Obtained)

Student : **Sourav Nandi**, MSc(Statistics) 1st year, IIT Kanpur

Advisor: **Dr. Ashish Ghosh**, Professor, MI Unit, ISI Kolkata

Period: 30 Nov 2017 to Current

**Brief Summary of Work Done**:

1. Implemented the Bayesian Classifier Algorithm (**from Scratch**), and used it in following Datasets from **UCI Machine Learning Repository**:

* **Wine Dataset**: 67.30% (Average Accuracy)
* Iris Dataset: 83.25%
* **Adult Dataset**: 58.275%
* Diabetes Dataset: 73.65%

1. Implemented the **Perceptron Algorithm (from Scratch)** in Python 3, and tested it on **Sonar Data Set**(Accuracy 76.923%) [https://http://archive.ics.uci.edu/ml/datasets/connectionist+bench+(sonar,+mines+vs.+rocks)]
2. Implemented a Classifier Algorithm( **from Scratch**) for a **Categorical Dataset**. Tested with **Flu DataSet** [https://www.youtube.com/watch?v=ZAfarappAO0 (Data Set given by Francisco Iacobelli)]. Also **mathematically justified the detailed flow of algorithm.**
3. Completed the MOOC Course “**Learning From Data**” offered by **CalTech** (Yaser S. Abu Mostafa) (<https://work.caltech.edu/telecourse.html>)
4. Implemented the Naïve Bayes Classifier Algorithm in **R** on:

* **Iris Dataset** (Accuracy 97.91%)
* **HouseVote 84 Dataset** (Accuracy 91.89%)
* Packages Used: e1071, mlbench

1. Bayesian Classifier Algorithm(From Scratch, Different from 1) in Python for **Pima Indian Diabetes Data Set** (Accuracy 74.92%). [https://archive.ics.uci.edu/ml/datasets/Pima+Indians+Diabetes]
2. Studied the following topics from “**Pattern recognition**” by Sergios Theodoridis and Konstantinos Koutroumbas and **used them to implement different machine learning algorithms**: Bayes Decision theory, different Distance Measures and Minimum Distance Classifier, MLE and MAP method of estimation, Maximum Entropy estimation, Least Square, MSE, The **Perceptron Algorithm**.
3. Implemented the following Classification Algorithms in **Python** on Iris Dataset (https://archive.ics.uci.edu/ml/datasets/iris):

* **Gaussian Naïve Baye**s Algorithm (Accuracy 98.33%)
* **Multinomial Naïve Bayes** Algorithm (Accuracy 96.67%)
* Library Used : Scikit Learn