

# Report

## Assignment 4 - Data Mining

---

### Group Members:

Sahil Sangwan : (2015A7PS0965H)  
Pranav Kumar Asthana : (2015A7PS0961H)  
Uttara Ravi : (2015A7PS0032H)  
Utkarsh Grover : (2017A7PS1428H)

---

### Preprocessing:

Naive bayes implementation did not require any preprocessing. A 80:20 train:test split was used.

For knn, the columns of the data were split to create a one hot encoded system which would be used to find the euclidean distance between different entries. A distance matrix is also created as part of the preprocessing part.

---

### Dataset Used:

The nursery dataset was used: <https://archive.ics.uci.edu/ml/datasets/nursery>

---

### Problem 1:

For the knn problem, a variety of k values were tested for. The best accuracy was noted for k = 14-18. The best one being for k=17.

**Accuracy:** 96.952

**Confusion Matrix: (Predicted- Actual|)**

	not_recom	priority	spec_prior	very_recom
not_recom	859	0	0	0
priority	0	818	21	0
spec_prior	0	12	812	0
very_recom	0	46	0	24

**Recall of not\_recom : 100.0**

**Precision of not\_recom : 100.0**

**Recall of priority : 97.49702026221692**

**Precision of priority : 93.37899543378995**

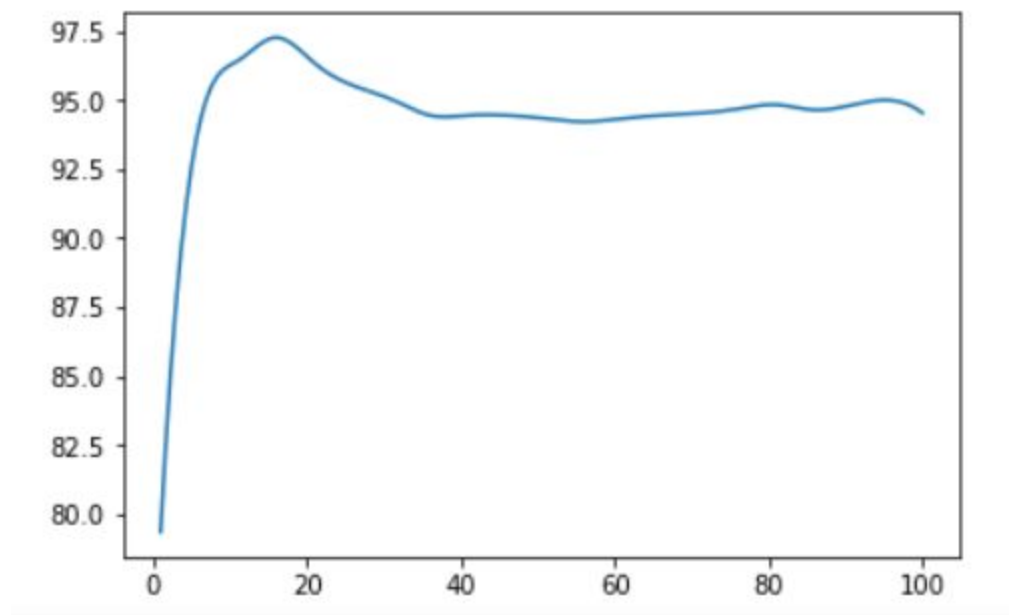
**Recall of spec\_prior : 98.54368932038835**

**Precision of spec\_prior : 97.47899159663865**

**Recall of very\_recom : 34.285714285714285**

**Precision of very\_recom : 100.0**

A plot of the k-value versus accuracy is given below which depict a peak at k = 14-18



---

## Problem 2:

A naive Bayes classifier was implemented in the second problem.

**Accuracy** = 70.910

**Confusion Matrix:**

	not_recom	priority	spec_prior	very_recom
not_recom	843	0	0	0
priority	0	498	379	0
spec_prior	0	308	497	0
very_recom	0	67	0	0

**Recall of not\_recom** : 100.0

**Precision of not\_recom** : 100.0

**Recall of priority** : 56.78449258836944

**Precision of priority** : 57.0446735395189

**Recall of spec\_prior** : 61.73913043478261

**Precision of spec\_prior** : 56.7351598173516

**Recall of very\_recom** : 0.0

**Precision of very\_recom** : 0.0