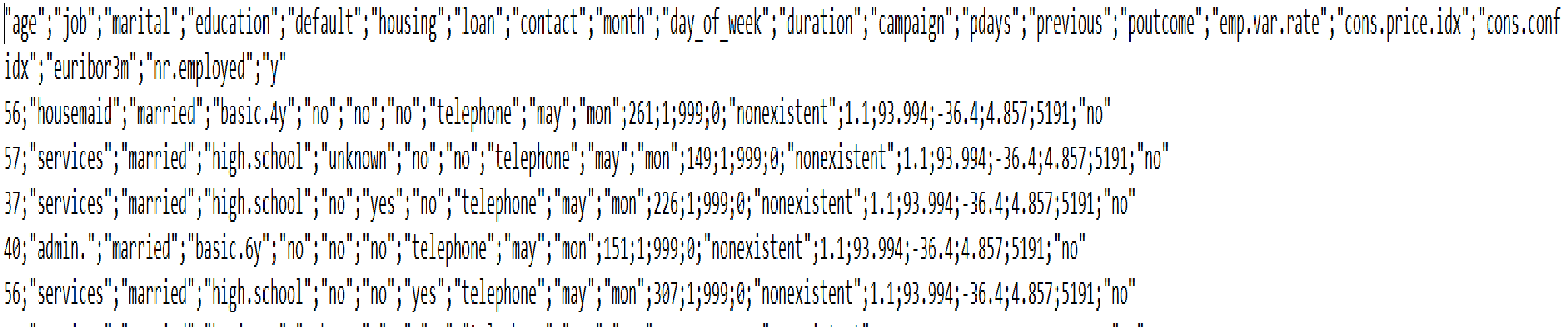
Final project presentation

By

Kiran Kumar Chilla

Dataset

- Dataset consists of 41,188 rows with 20 attributes
- Goal is to predict whether client will subscribe term deposit or not
- Field benefitting with the analysis : Banking institutions



```
"age";"job";"marital";"education";"default";"housing";"loan";"contact";"month";"day_of_week";"duration";"campaign";"pdays";"previous";"poutcome";"emp.var.rate";"cons.price.idx";"cons.conf.
idx";"euribor3m";"nr.employed";"y"
56;"housemaid";"married";"basic.4y";"no";"no";"no";"telephone";"may";"mon";261;1;999;0;"nonexistent";1.1;93.994;-36.4;4.857;5191;"no"
57;"services";"married";"high.school";"unknown";"no";"no";"telephone";"may";"mon";149;1;999;0;"nonexistent";1.1;93.994;-36.4;4.857;5191;"no"
37;"services";"married";"high.school";"no";"yes";"no";"telephone";"may";"mon";226;1;999;0;"nonexistent";1.1;93.994;-36.4;4.857;5191;"no"
40;"admin.";"married";"basic.6y";"no";"no";"no";"telephone";"may";"mon";151;1;999;0;"nonexistent";1.1;93.994;-36.4;4.857;5191;"no"
56;"services";"married";"high.school";"no";"no";"yes";"telephone";"may";"mon";307;1;999;0;"nonexistent";1.1;93.994;-36.4;4.857;5191;"no"
```

Dataset

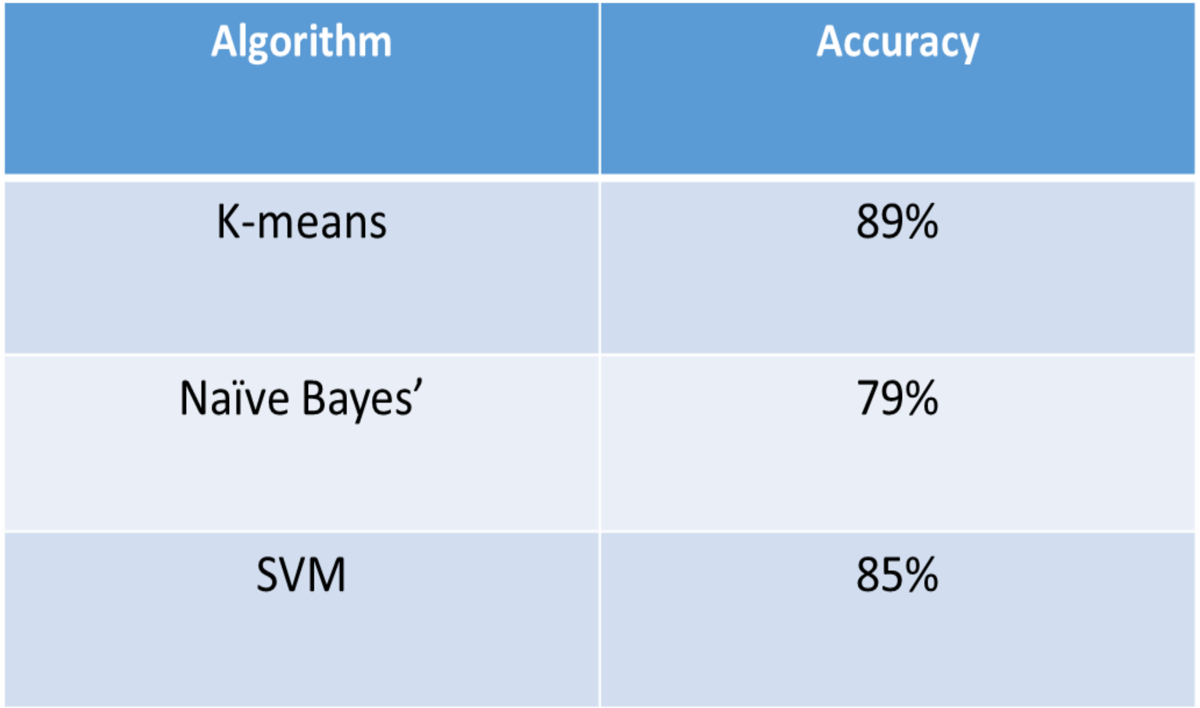
- ➢The target variable will be 1 or 0
- ➢It has 10 categorical and 10 numeric variables
- ➢3 types of predictor variables
  1. *Demographic data*(age, job, marital, default, education, housing, loan)
  2. *Data related to previous contact*(contact, month, day of week, duration, campaign, pdays, previous, poutcome)
  3. *Data collected during the contact made* (employment rate, consumer price index, consumer confidence index, euribor3m indicator, and number of employees)

Data cleaning and Correlation

- Default column is removed since 99% of the default value is ‘no’
- 96% of the pvalues consists ‘999’ as data, so it is removed
- Dataset will be reduced to [37050,17]
- Correlation between input and target variable is calculated
- This correlation coefficient is used in choosing inputs with high correlation coefficient over all other inputs

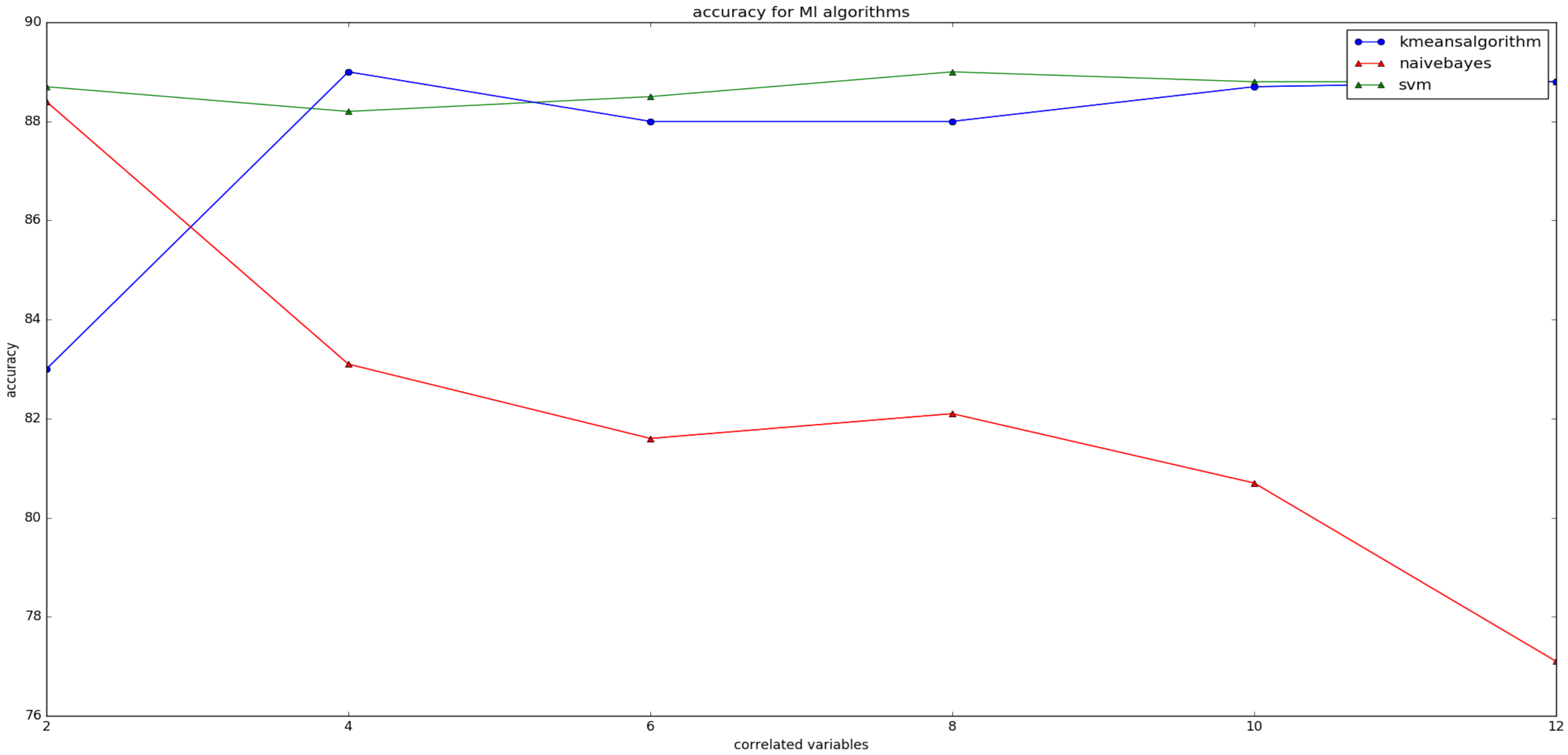
Machine learning algorithms

- Kmeans
- Naive bayes
- Svm
- Neural networks



Algorithm	Accuracy
K-means	89%
Naïve Bayes’	79%
SVM	85%

Selecting correlated variables


accuracy for Ml algorithms
kmeansalgorithm
naivebayes
svm
accuracy
correlated variables