\* Gaussian Naive Bayes (GaussianNB)-Simple and quicker to converge

\* Decision Trees - Easy to interpret and non-parametric

\* Support Vector Machines (SVM) - High accuracy and good guarantees regarding over fitting.

Gaussian Naive Bayes (GaussianNB)

- Describe one real-world application in industry where the model can be applied.

\* Bag of words and Text classification problems.

\* Spam classification.

- What are the strengths of the model; when does it perform well?

\* Simple and highly scalable

\* Converges easily when the conditionally independent assumption holds.

\*Able to handle many features

\*Naive Bayes can be used for Binary and Multiclass classification.

- What are the weaknesses of the model; when does it perform poorly?

\*It considers all the features to be unrelated, so it cannot learn the relationship between features.

\*It will have bias on very small data sets.

- What makes this model a good candidate for the problem, given what you know about the data?

\*This is simple data set where the Naïve Bayes can produce a good convergence and good Training time.

Decision Trees

- Describe one real-world application in industry where the model can be applied.

\*Fraudulent financial statement detection

\* Speech Recognition technologies

- What are the strengths of the model; when does it perform well?

\* Easy to interpret and explain

\* They are non-parametric so do not have to worry about non-linearly separable

\*Good for datasets with missing value attributes

\*Fast and Scalable.

- What are the weaknesses of the model; when does it perform poorly?

\* It will easily over fit

\*It doesn’t support online learning so you will need to rebuild as new examples come

- What makes this model a good candidate for the problem, given what you know about the data?

\*It easy and fast to build. The information gain can find important attributes considering the limited domain knowledge.

SVM

- Describe one real-world application in industry where the model can be applied.

\* Bioinformatics (Protein classification, Cancer classification).

\* Hand-written character recognition.

\* image classification problems.

- What are the strengths of the model; when does it perform well?

\* High accuracy and not easily over fitted.

\* Data doesn’t need to be linearly separable if an appropriate Kernel method is used.

\*Captures complex relationships in data.

- What are the weaknesses of the model; when does it perform poorly?

\* Memory intensive.

\* Difficult to tune and run due to too many parameters.

\* It is sensitive to noise.

- What makes this model a good candidate for the problem, given what you know about the data?

\*SVM is a good classifying model and it works well with high number of features as we have 103 features after hot encoding.

References Used -http://blog.echen.me/2011/04/27/choosing-a-machine-learning-classifier/

-http://dataaspirant.com/2017/02/06/naive-bayes-classifier-machine-learning/