**Experimantal Environment**

OS – Windows 8.1, Ubuntu

Softwares used – 1. Jflap : for verifcation of reglar language

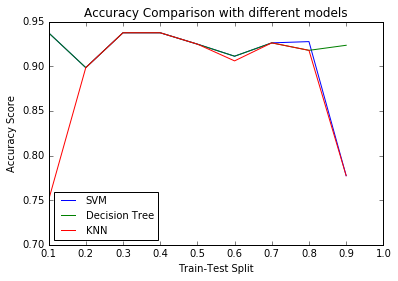
2.Jupyter notebook : for python scripting

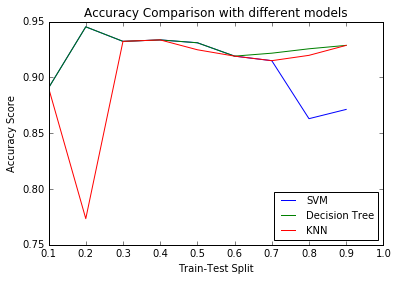
3. Sciitkit Learn : Machine Learning Classifiers library

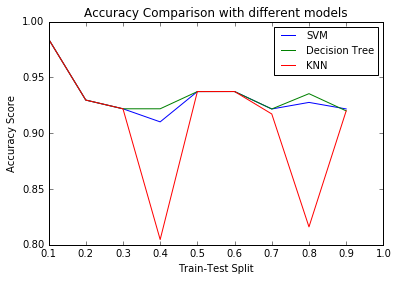
4. Eclipse IDE : for writing parser in JAVA

**Testing data pie chart**

**Classifier accuracy plot**





****

**Why FSM**

1.As we have seen above the accuracy score of different ML models vary depending upon the splitting of available data.

Also for different data size the same splitting will result in different accuracy scores.

By using FSM for constructing the model we will eliminate this shortcoming.

2. Also it is not possible to train models with different number of features(which highly observe in tweets, i.e. tweets of different lengths)

Hence the bag of word will differ for insance to instance. To overcome this situation we have to extract fixed number of features and have to neglect the other ones.

3. By using FSM, we will be able to retain the dependency between different features but that is not possible in ML training models.

Like coming of “U\_NAME after @ “ and “TREND after #” can not be checked in ML training models.

And this shortcomings highly suggest to use FSM instead of ML training.

**RE for final DFA**

Q0=lambda

Q1=Q0.RT+Q4.RT

=RT+Q4.RT

Q2=Q0.@+Q1.@+Q3.@+Q4.@

Q3=Q2.U\_NAME

Q4=ENG+T+Q3.(ENG+T)+Q4.(ENG+T) +Q6.TREND

Q6=Q0.#+Q3.#+Q4. #

=# +Q3.# +Q4.#

To find the RE for state Q4 which is a final state we expressed every other term in terms of Q4, so that we can use Arden’s theorem

Q0 and Q1 need no more modification.

Q2=@+RT.@+Q4(RT.@+@) +Q2.U\_NAME.@

= (@+RT.@+Q4. (RT.@+@)). (U\_NAME.@) \* [using Arden’s theorem]

Q3=Q2.U\_NAME= (@+RT.@+Q4. (RT.@+@)). (U\_NAME.@) \*. U\_NAME

Q4 don’t need modification.

Q6=# +Q3.# +Q4.#

We get

Now putting values in terms of Q4 IN

Q4=ENG+T+Q3.(ENG+T)+Q4.(ENG+T) +Q6.TREND

We get

Q4= [ ENG +

T +

#. TREND+

(@+RT.@) (U\_NAME.@) \*.U\_NAME.(ENG + #. TREND+T)

] +

Q4. [ ENG +

T +

#. TREND +

(@+RT.@) (U\_NAME.@) \*.U\_NAME.(ENG + #. TREND+T) ]

Now using Arden’s theorem, we get RE for Q4 and hence for the FSM as follows

Q4= [ ENG + #. TREND+ (@+RT.@) (U\_NAME.@) \*.U\_NAME.(ENG + #. TREND+T)]. [ ENG + (@+RT.@) (U\_NAME. @) \*.U\_NAME.(ENG + #. TREND+T) + #. TREND] \*

=[ ENG + #. TREND+ (@+RT.@) (U\_NAME.@) \*.U\_NAME.(ENG + #. TREND+T)] +