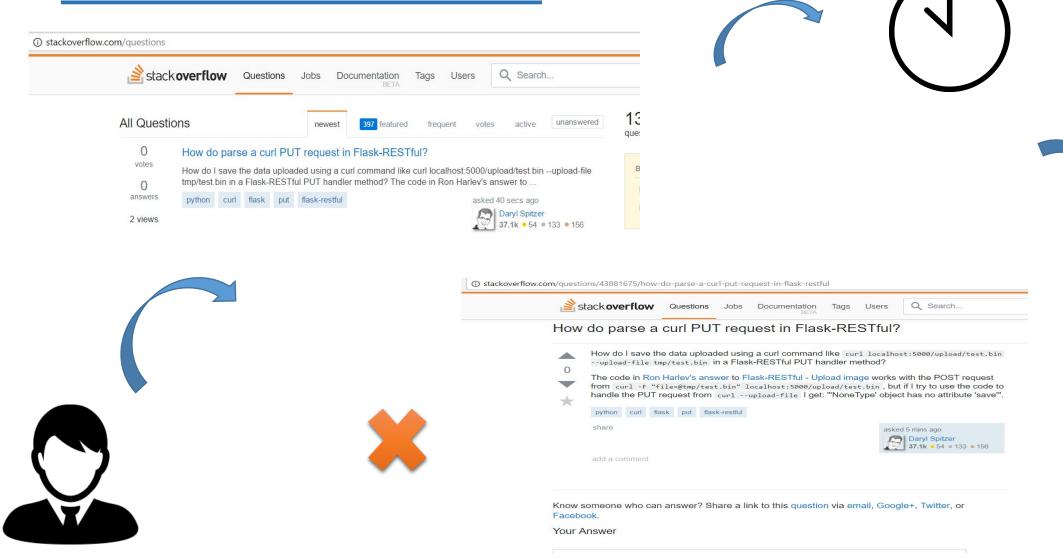
Stackoverflow – Time until Answer

Team – 4

Gaurang Mhatre
Parth Upadhyay
Amruta Dhondage

Problem Statement:



Tools / Technologies used:











Numpy & SciPy Essential Training

Dataset used:

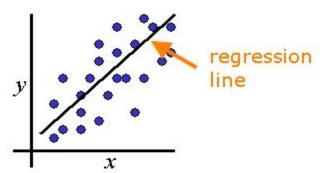
https://www.ics.uci.edu/~duboisc/stackoverflow/

Records: 3162491

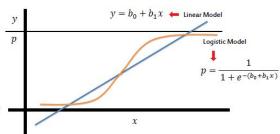
. S. S. S.				-		- I				1.7	
	qid	i	qs	qt	tags	qvc	qac	aid j	as		at
1	563355	62701	0	1.24E+09	php,error,gd,image-processing	220	2	563372	67183	2	1.24E+09
2	563355	62701	0	1.24E+09	php,error,gd,image-processing	220	2	563374	66554	0	1.24E+09
3	563356	15842	10	1.24E+09	lisp,scheme,subjective,clojure	1047	16	563358	15842	3	1.24E+09
4	563356	15842	10	1.24E+09	lisp,scheme,subjective,clojure	1047	16	563413	893	18	1.24E+09
5	563356	15842	10	1.24E+09	lisp,scheme,subjective,clojure	1047	16	563454	11649	4	1.24E+09
6	563356	15842	10	1.24E+09	lisp,scheme,subjective,clojure	1047	16	563472	50742	6	1.24E+09

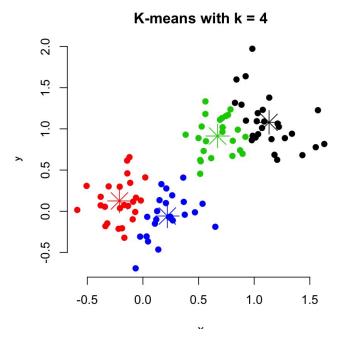
Approach:

Progressive data: Regression

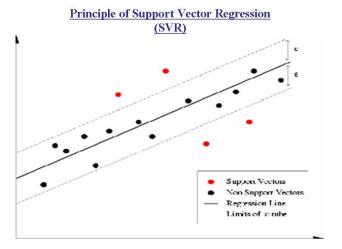


Linear Regression





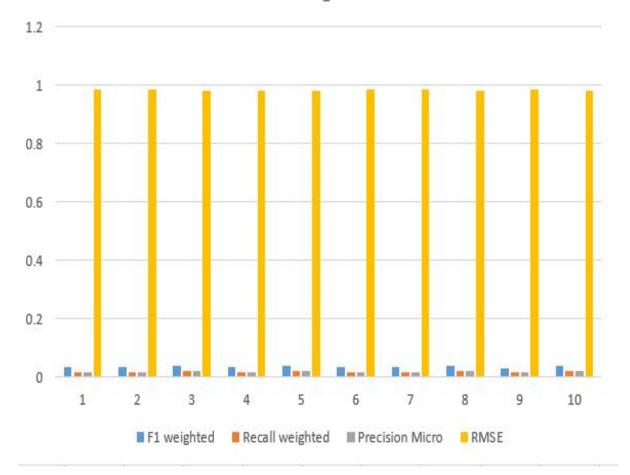
Clustering Regression



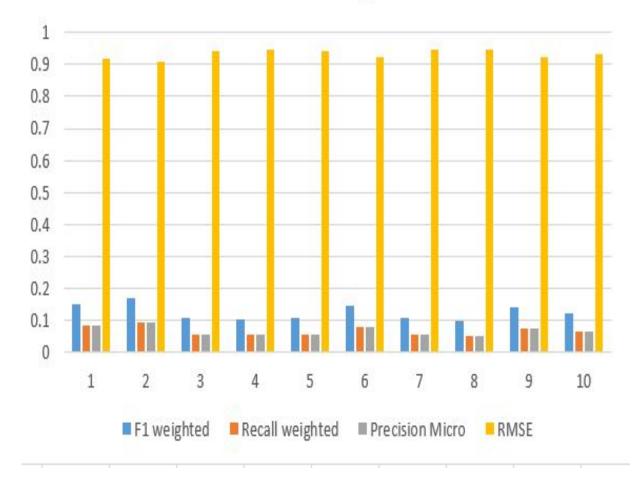
Logistic Regression

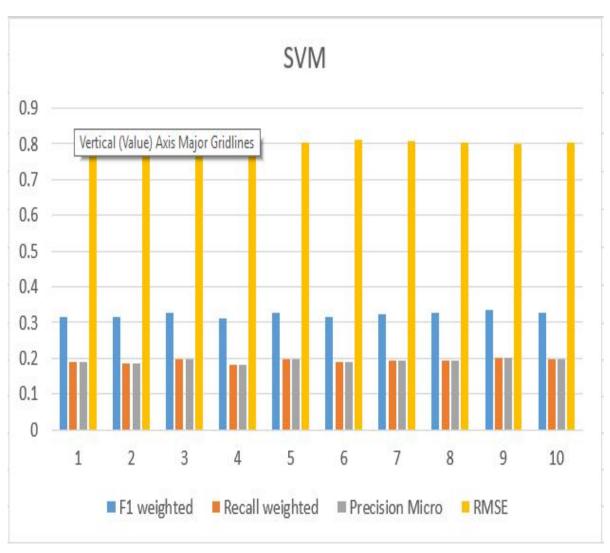
SVM Regression

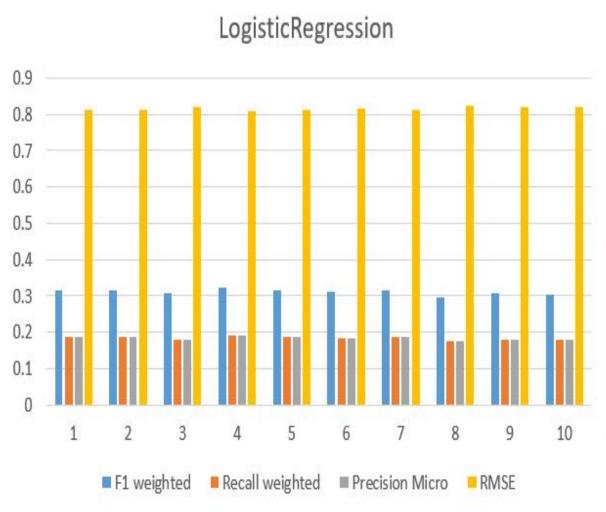




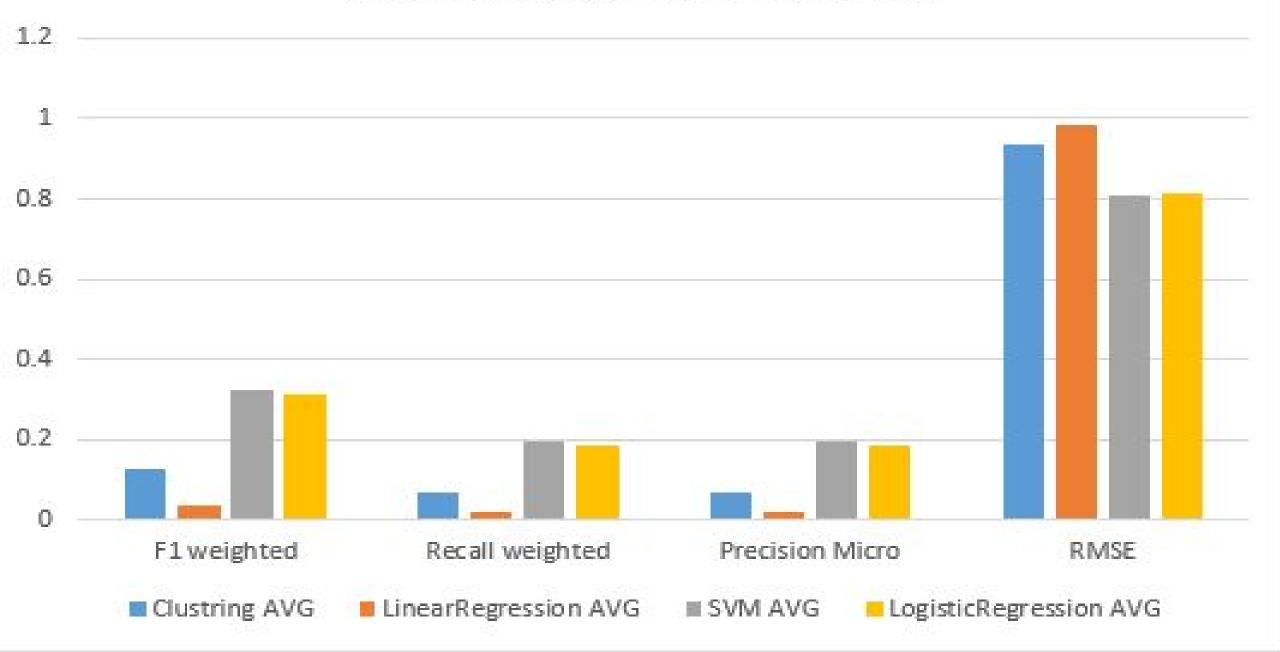
Clustering







Average Results for 10 iterations



Conclusion

Regression Data mining algorithm is used for predicting time. Clustering approach is experimented for prediction. Questions??