OSSASSIST

Deepthi Venkitaramanan Rudra Purohit Nitin Kodial Biswajyoti Pal

Georgia Tech

OssAssist-Release Cycle Prediction

Statistics

BELOW ARE THE PREDICTION OF NUMBER OF CONTRIBUTORS FOR THE CURRENT TERM ALONG WITH THE CURRENT TERMS' ISSUES AND PULL REQUESTS(A TERM IS A PERIOD OF 45 DAYS)

1036 1069 76 339

Recent News   month week day

< today   **March 2016**   >

Sun	Mon	Tue	Wed	Thu	Fri	Sat
28	29	1	2	3	4	5
6	7	8	9	10	11	12
13	14	15	16	17	18	19
20	21	22	23	24	25	26
27	28	29	30	31	1	2
3	4	5	6	7	8	9
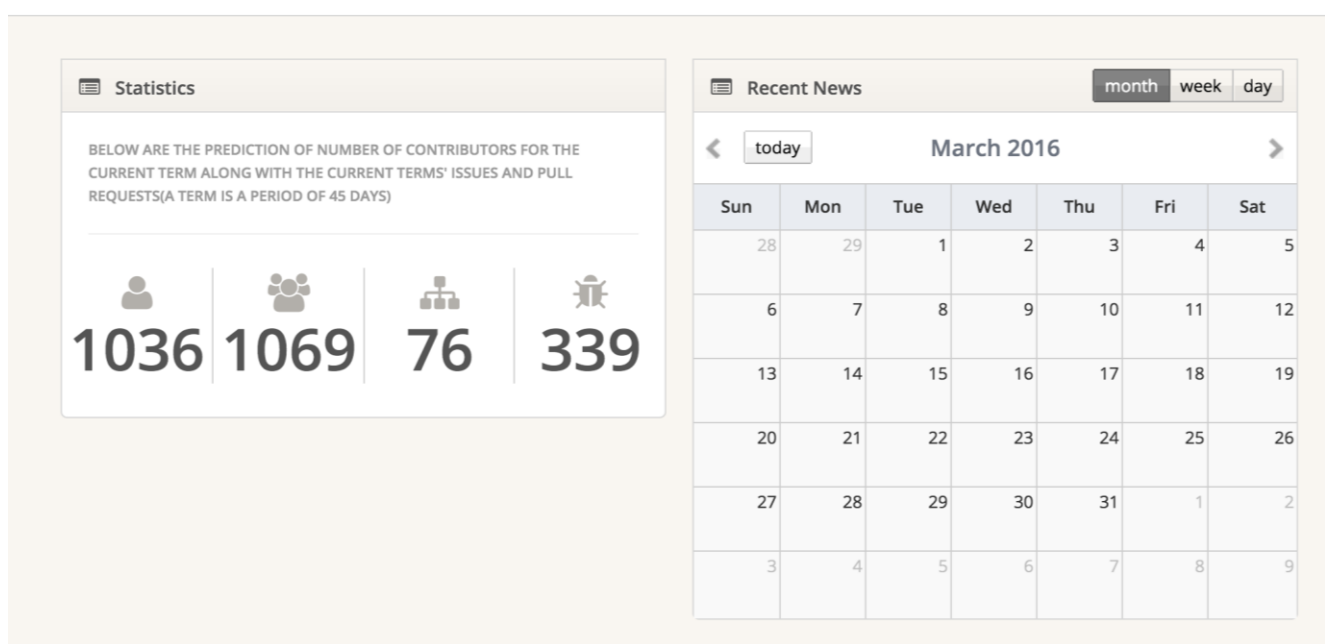
- **Summary**: OssAssist enables OSS organizations predict the number of contributors and visualize key performance indicators of their projects.
- This system shall be an important tool for planning development cycles and releases.
- We used the GitHub API to obtain data (15GB) corresponding to 'Mozilla' and applied machine learning algorithms on the extracted features and attributes to predict the number of contributors.

Approach 1

- **Regression- Linear**
- Repository Data for current term predicts contributor count for next term:
- Term 1 [Issues count, Comments Count, Pull Request Count, Contributors Count, Issue Comments count, Tags Count, Commits count] ---> Term 2 [Contributors count]

Approach 2

- **Regression- Multiple Linear**
- Maximum Fit based on R square[coefficient of determination] and Mean Squared Error
- For predicting Nth Term contributors count, the attributes of the repositories and contributors count are considered for all N-1 time slices.

Approach 3

- **Classification based on single term data**
- Term 1 [Issues count, Comments Count, Pull Request Count, Contributors Count, Issue Comments count, Tags Count, Commits count] ---> Term 2 [Contributors count]
- However, Term 2 contributors count here is a label(range of no of contributors) split based on Jenks natural split.

Approach 4

- **Classification based on all time slice data**
  For predicting Nth Term contributors count, the attributes of the repositories and contributors count are considered for all N-1 time slices.
- However, Term 2 contributors count here is a label(range of no of contributors) split based on Jenks natural split.
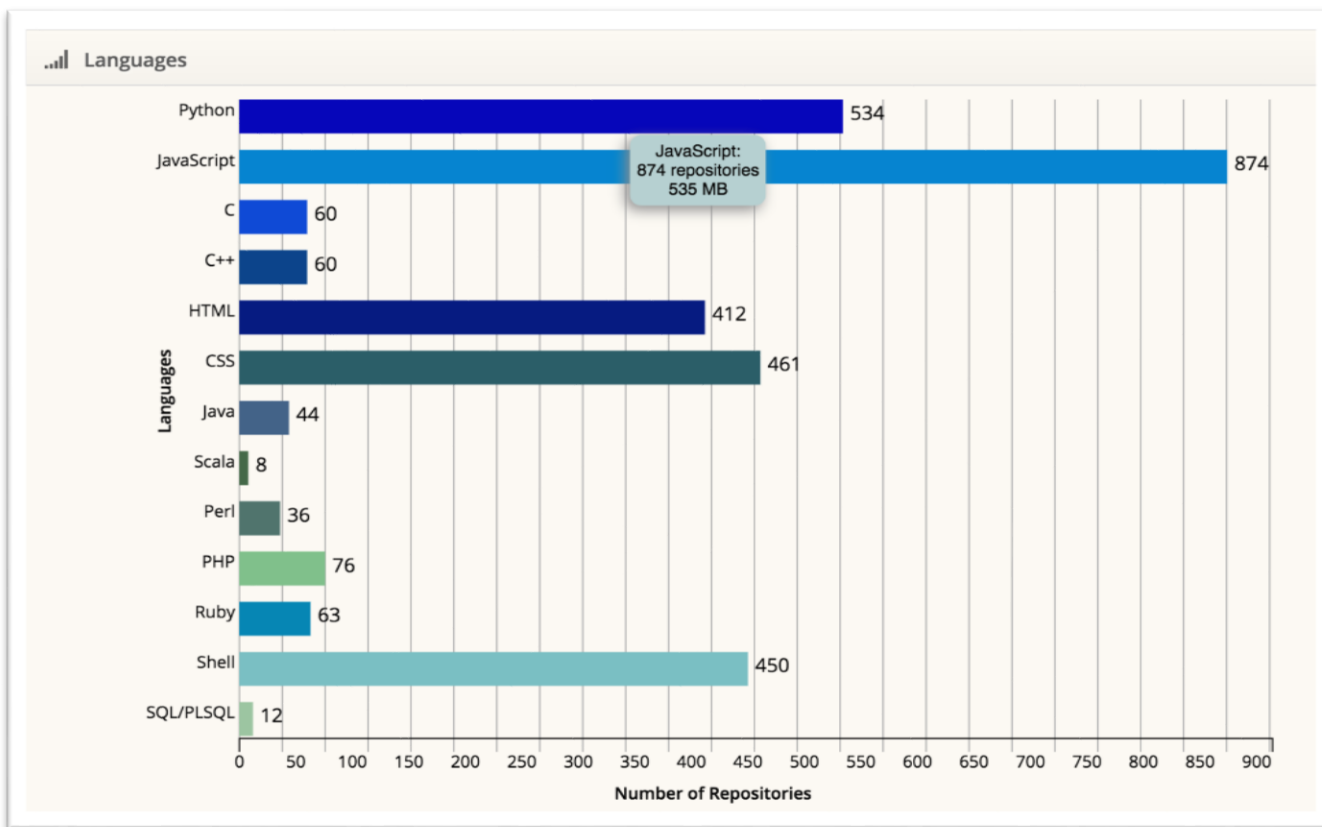
Experiments and Results

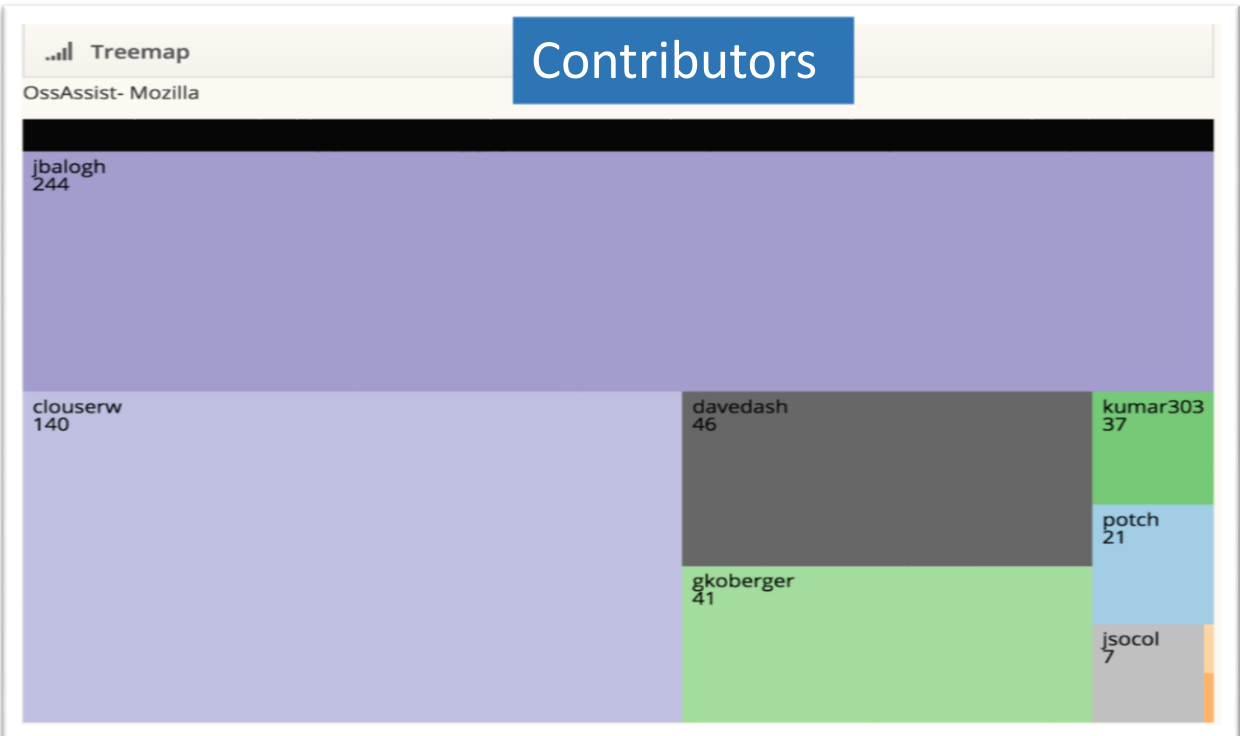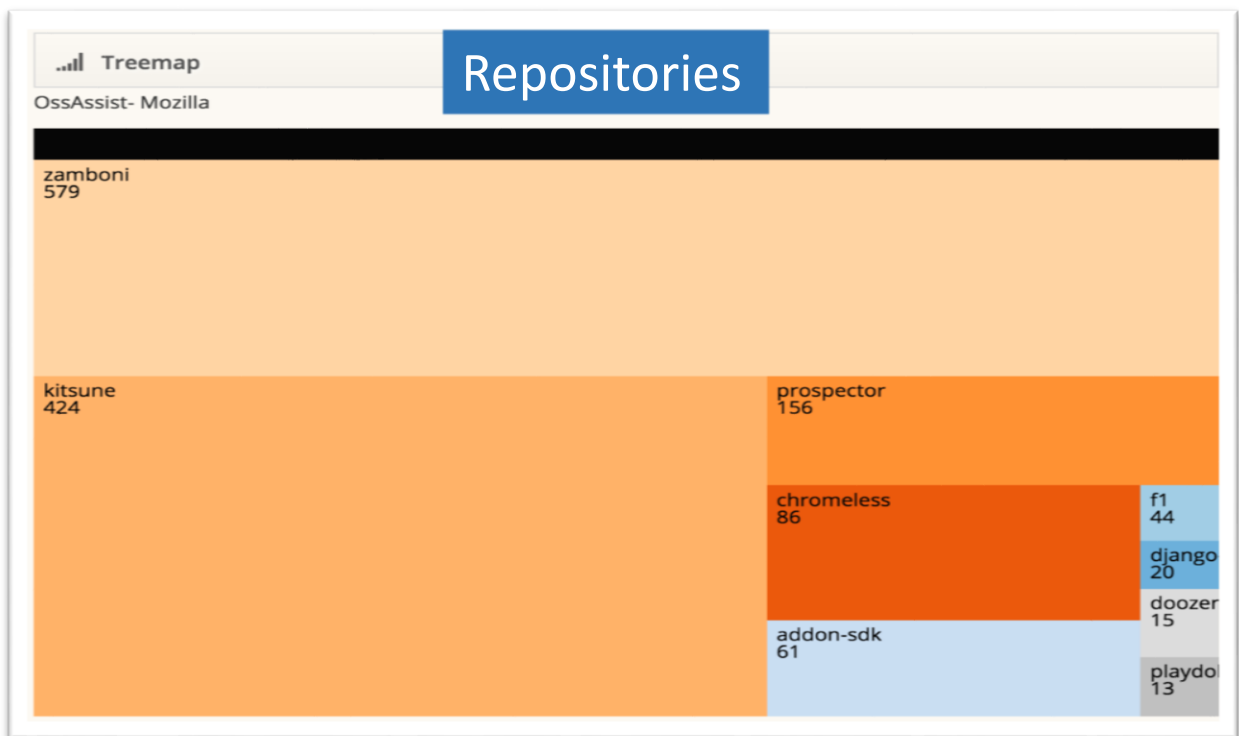| Regression Techniques	Regression (1 time slice): R2 Score	Regression (n time slices): R2 Score
Linear Regression	0.63	0.72
ElasticNet	0.64	0.74
Lasso	0.64	0.74

Regression Techniques	Regression (1 time slice): Mean Squared Error	Regression (n time slices): Mean Squared Error
Linear Regression	20.31	18.43
ElasticNet	19.67	17.60
Lasso	19.67	17.53

Popular Languages in the Organization

Languages

Language	Number of Repositories
Python	534
JavaScript	874 (JavaScript: 874 repositories 535 MB)
C	60
C++	60
HTML	412
CSS	461
Java	44
Scala	8
Perl	36
PHP	76
Ruby	63
Shell	450
SQL/PLSQL	12

Drill Down for a Repository

Repositories

Treemap
OssAssist- Mozilla

zamboni 579
kitsune 424
prospector 156
chromeless 86
f1 44
django 20
doozer 15
addon-sdk 61
playdo 13

Contributors

Treemap
OssAssist- Mozilla

jbalogh 244
clouserw 140
davedash 46
kumar303 37
gkoberger 41
potch 21
isocol 7

Key Technologies & Components