

Predicting **Visitor Engagement** For Blogs

By
Arya Das,
Karthik Keertipati &
Suleiman Ali Shakir

Background

More than **440M** blogs in the world



Blogs are the most used sites for **information**



Professional Bloggers make **\$20k-\$150k** a year

Problem

Blogs don't ask for
personal user data



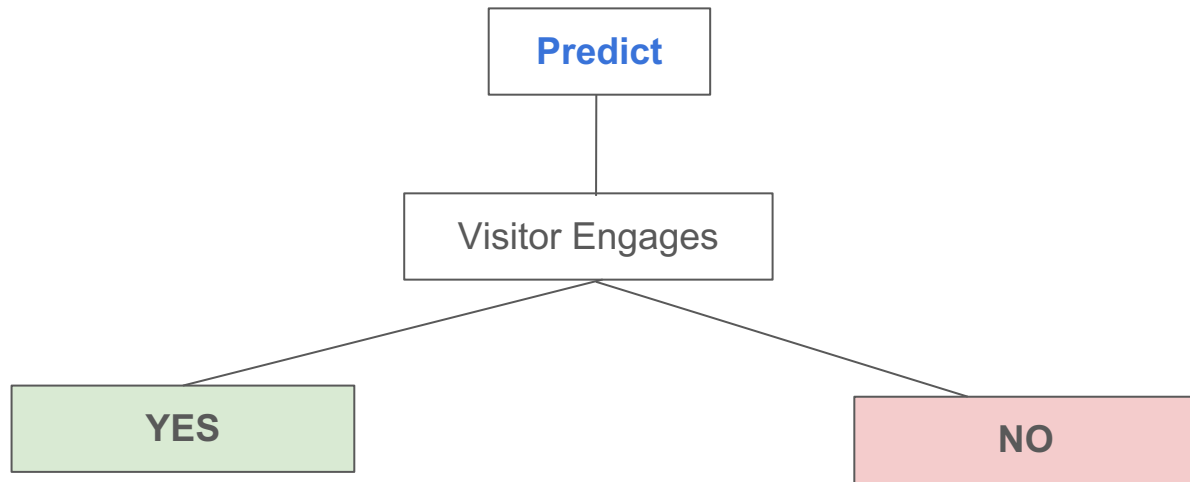
Cannot understand who
the **best visitors** are



Blogs have extremely
high **Bounce Rate**



Goal Definition



- Email sign-up prompts
- Special deals/coupons
- Recommend articles

- Remove ads to improve UX for better content engagement

Literature Review

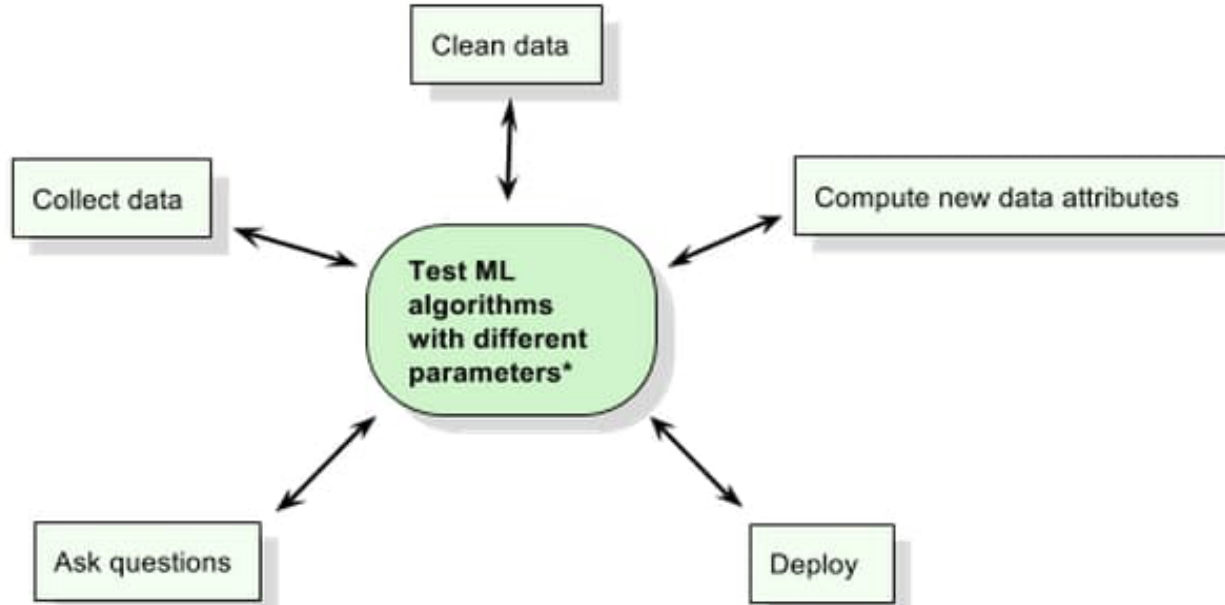
1. Machine Learning on Google Analytics
by Fabrice Hong
2. Predicting Bounce Rates in Sponsored Search Advertisements
by D. Sculley, Robert Malkin, Sugato Basu, Roberto J. Bayardo
3. Critical Regression Analysis of Real Time Industrial Web Data Set using Data Mining Tool
by Shruthi Kohli and Ankit Gupta
4. Metrics to Measure Website Effectiveness
by Weissenburger, Andrew
5. Leveraging Google
by Malisa Anderson-Strait
6. How SOASTA & Google used ML to predict Bounce Rate & Conversions
by Tammy Everts

Learning Outcomes

- Selecting which Attributes to use
- How to distribute positive and negative class across the dataset
- Which Algorithms to use
- How to evaluate and which evaluation metrics to use



Methodology



* Experimentation software. "Weka" in our case

Approach

Data Exploration

Data Source



Exported data from Google Analytics

Minute Index	Landing Page	Country	Operating System	Browser	Session Duration	Bounce Rate
1. 000888	/android-pagination-tutorial-getting-started-recyclerview/	South Korea	Windows	Chrome	01:15:06 (1.31%)	0.00%
2. 009367	/horizontal-scrolling-lists-mobile-best-practices/	United States	Windows	Chrome	01:15:04 (1.21%)	0.00%
3. 002094	/bottom-navigation-bar-android-tutorial/	India	Windows	Chrome	01:11:13 (1.19%)	0.00%
4. 006323	/easy-navigation-drawer-with-design-support-library/	Indonesia	Windows	Opera	01:06:09 (1.07%)	0.00%
5. 002405	/android-pagination-tutorial-getting-started-recyclerview/	India	Linux	Firefox	01:02:06 (1.00%)	0.00%
6. 005565	/material-design-for-tablets/	Rwanda	Windows	Firefox	00:51:39 (0.83%)	0.00%
7. 008568	/parallax-scrolling-tabs-design-support-library/	Brazil	Windows	Firefox	00:50:34 (0.82%)	0.00%
8. 002208	/google-chrome-custom-tabs-android-tutorial/	Iran	Android	Chrome	00:47:22 (0.76%)	0.00%
9. 004764	/android-pagination-error-handling/	South Korea	Windows	Chrome	00:45:34 (0.74%)	0.00%
10. 010032	/onboarding-android-viewpager-google-way/	Azerbaijan	Windows	Chrome	00:45:03 (0.73%)	0.00%

Minute Index	Landing Page	Source	Device Category	Session Duration	Bounce Rate
1. 000888	/android-pagination-tutorial-getting-started-recyclerview/	gust.github.com	desktop	01:15:06 (1.31%)	0.00%
2. 009367	/horizontal-scrolling-lists-mobile-best-practices/	google	desktop	01:15:04 (1.21%)	0.00%
3. 002094	/bottom-navigation-bar-android-tutorial/	google	desktop	01:11:13 (1.19%)	0.00%
4. 006323	/easy-navigation-drawer-with-design-support-library/	duckduckgo.com	desktop	01:06:09 (1.07%)	0.00%
5. 002405	/android-pagination-tutorial-getting-started-recyclerview/	google	desktop	01:02:06 (1.00%)	0.00%
6. 005565	/material-design-for-tablets/	google	desktop	00:51:39 (0.83%)	0.00%
7. 008568	/parallax-scrolling-tabs-design-support-library/	google	desktop	00:50:34 (0.82%)	0.00%
8. 002208	/google-chrome-custom-tabs-android-tutorial/	google	mobile	00:47:22 (0.76%)	0.00%
9. 004764	/android-pagination-error-handling/	gust.github.com	desktop	00:45:34 (0.74%)	0.00%
10. 010032	/onboarding-android-viewpager-google-way/	(direct)	desktop	00:45:03 (0.73%)	0.00%

Data Exploration

- 2 Data Files of different attributes, taken over **2 years**.

FILE 1

47013 instances

ATTRIBUTES :

- minute index
- landing page
- country
- operating system
- Browser
- session duration
- bounce rate

FILE 2

40480 instances

ATTRIBUTES :

- minute index
- landing page
- Source
- device category
- bounces

Data Preparation

Tools :



Process :

Read

JOIN

Derive Classes

- After Joining - 40439 instances

Data Sample

Landing Page	Country	Operating System	Browser	Session Duration	Bounce Rate	Source	Device Category	clas
/how-to-design-ap	United States	Macintosh	Chrome	0	1	youtube.com	desktop	0
/android-paginat	Ukraine	Windows	Chrome	0	1	google	desktop	0
/pinterest-masonr	Spain	Windows	Internet Expl	0	1	(direct)	desktop	0
/material-design-ta	India	Windows	Chrome	6632	0.5	google	desktop	1
/material-design-ta	India	Windows	Chrome	3362	0.5	google	desktop	1
/parallax-scrolling-	India	Windows	Chrome	3338	0.5	google	desktop	1
/bottom-navigation	India	Windows	Chrome	3087	0.5	google	desktop	1

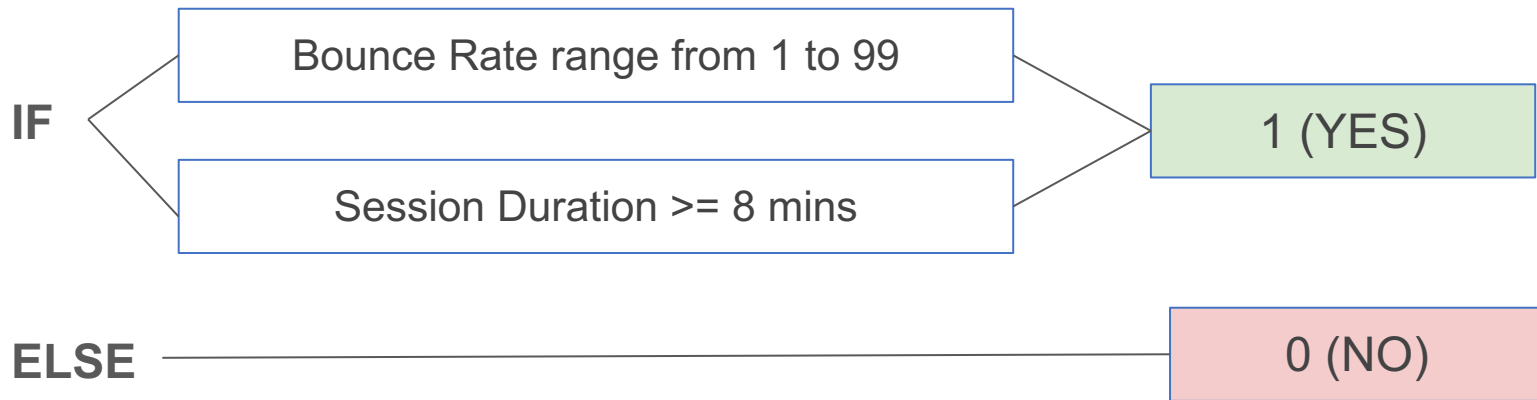
ARFF

```
1 @relation final
2
3 @attribute 'Landing Page' {/circular-reveal-effect-like-whatsapp-in-android/,/pinterest-masonry-layout-staggered-grid/,/create-gmail-
style-list-in-android/,/android-architecture-components-tutorial-room-livedata-viewmodel/,/toolbar-animation-with-android-design-
support-library/,/horizontal-scrolling-lists-mobile-best-practices/,/bottom-navigation-bar-android-tutorial/,/material-design-tabs-
with-android-design-support-library/,/android-pagination-tutorial-getting-started-recyclerview/,/android-pagination-recyclerview-
tutorial-api-retrofit-gson/,/runtime-permissions-android-marshmallow/,/bottom-sheet-android-design-support-library/,/android-sqlite-
database-tutorial-sugar-orm/,/image-gallery-app-android-studio-1-4-glide/,/parallax-scrolling-tabs-design-support-library/,/android-
pagination-error-handling/,/using-bottom-navigation-view-android-design-support-library/,/stop-using-loading-spinner-theres-something-
better/,/techniques-to-display-text-overlay-background-images/,/material-design-snackbar/,/onboarding-android-viewpager-google-
way/,/android-o-tutorial-custom-font-resources/,/create-a-card-toolbar/,/top-10-material-design-libraries-for-android/,/how-to-design-
app-mockups-like-a-pro/,/daynight-theme-android-tutorial-example/,/recyclerview-adapter-android-made-fast-easy/,/material-design-for-
```

Data Preparation

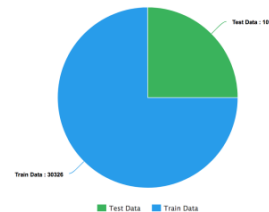
How did we define Classes?

Avg. Reading Time for this blog's articles is **at least 8 Mins**

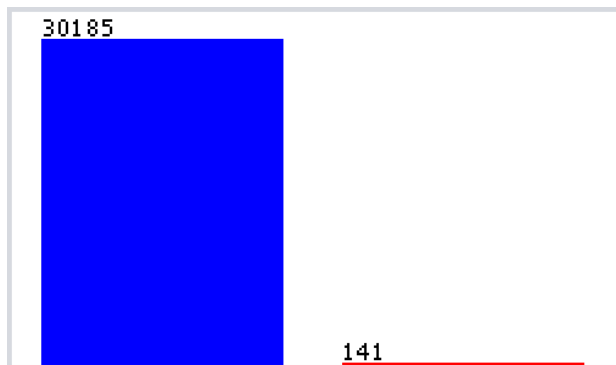


Data Preparation

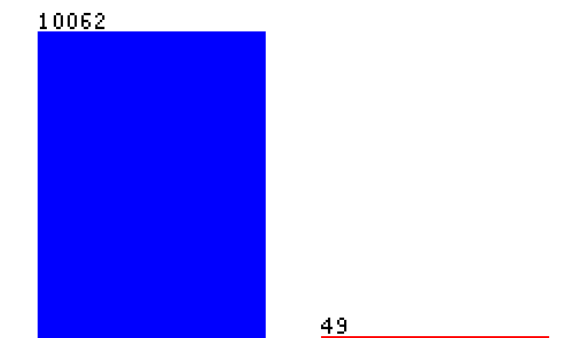
75% of data for training and 25% of data for testing



Training Data Classes



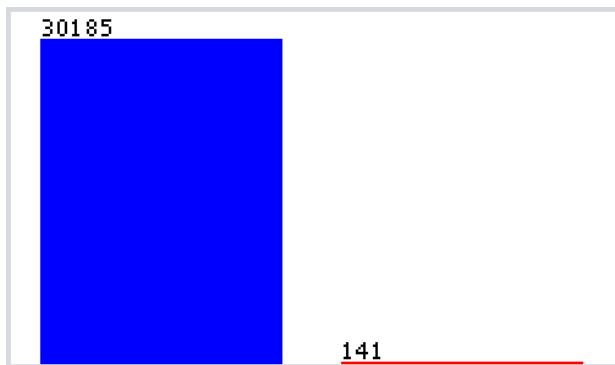
Testing Data Classes



Modelling

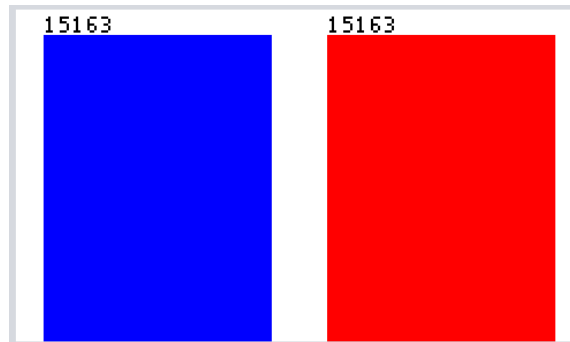
Before:

- Positive class: 141
- Negative class: 30185



ClassBalancer :

- Positive class: 15163
- Negative class: 15163



Modelling

- Algorithms :

- C4.5
- Naive Bayes
- Random Tree
- Random Forest
- OneR

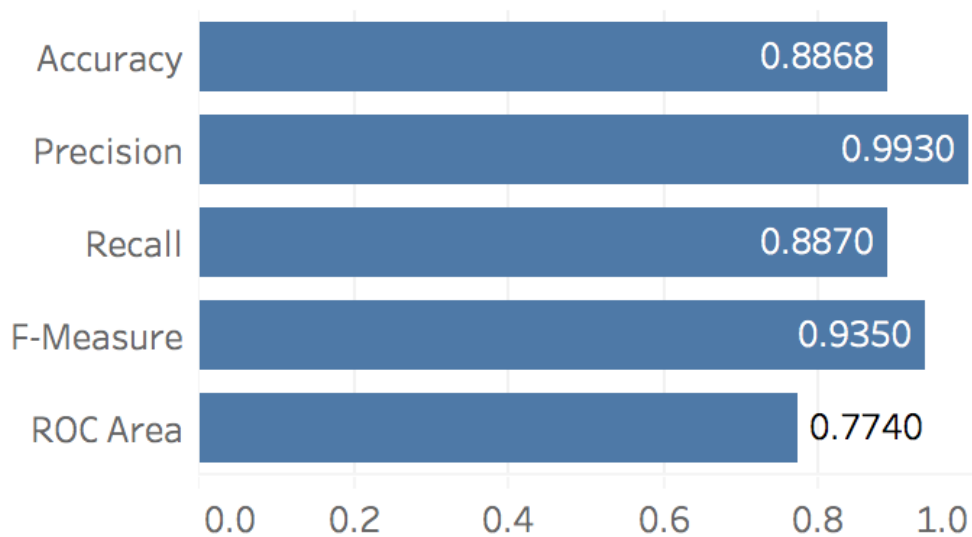
- Evaluation Metrics:

- Accuracy
- Precision
- Recall
- F-measure
- Area under ROC curve



Results

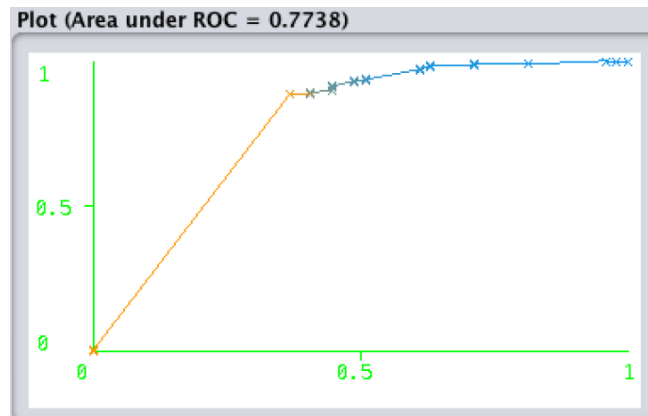
C 4.5 Algorithm (J48)



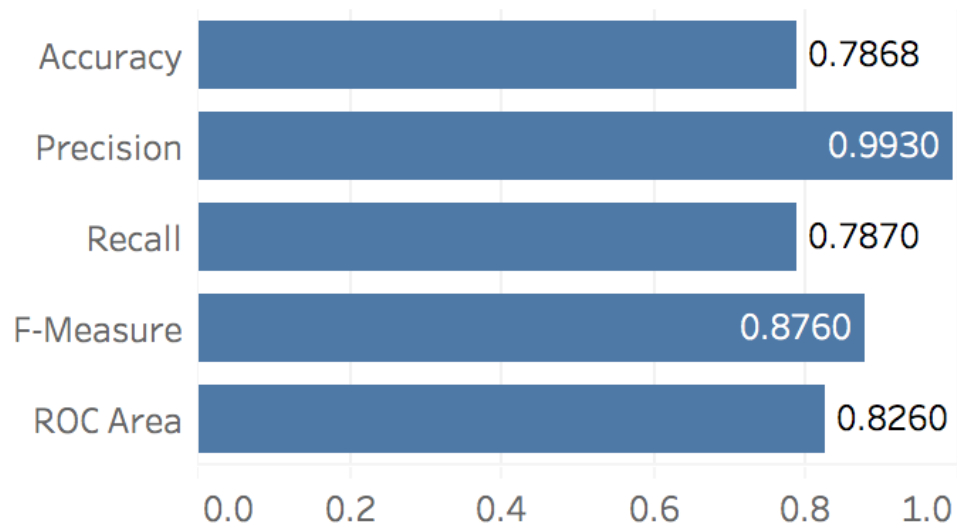
Confusion Matrix

0(no)	1(yes)	
8937	1125	0(no)
20	29	1(yes)

ROC Curve



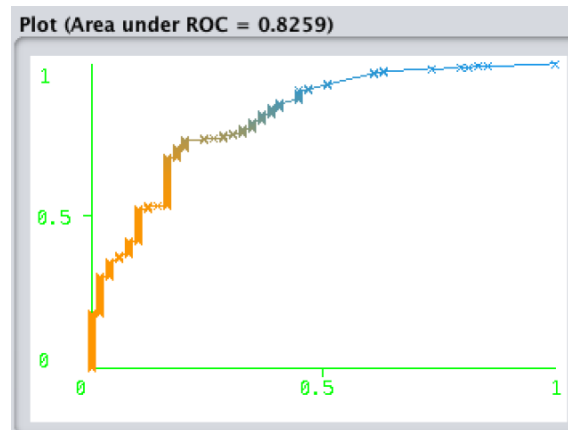
Naive Bayes



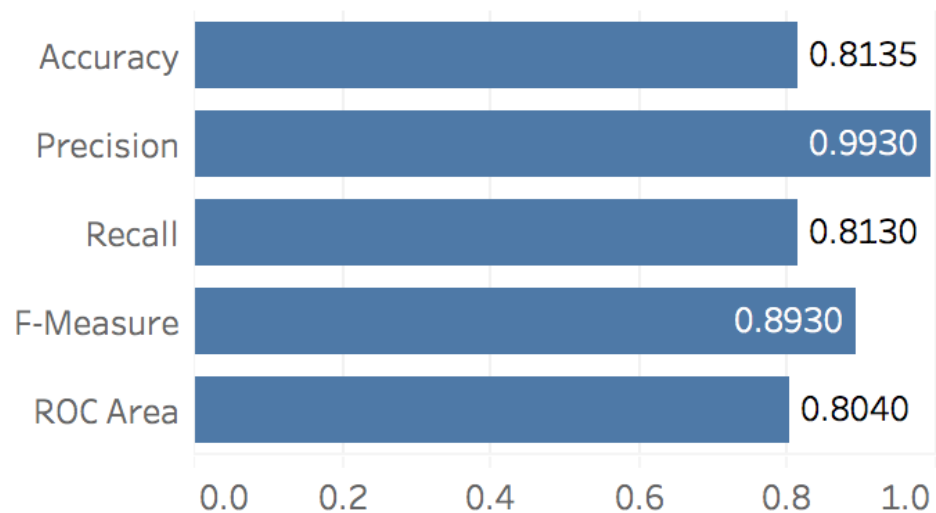
Confusion Matrix

0(no)	1(yes)	
7923	2139	0(no)
17	32	1(yes)

ROC Curve



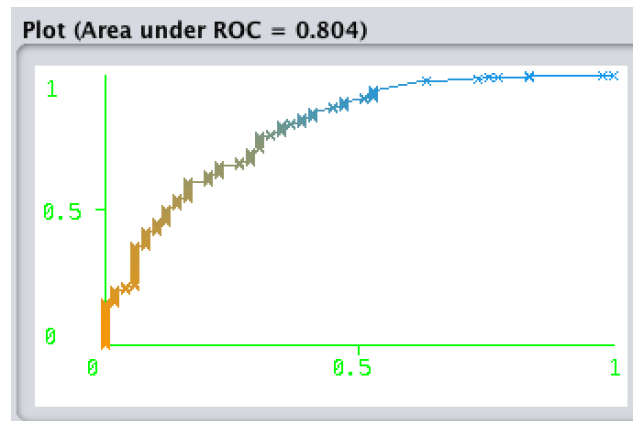
Random Forest



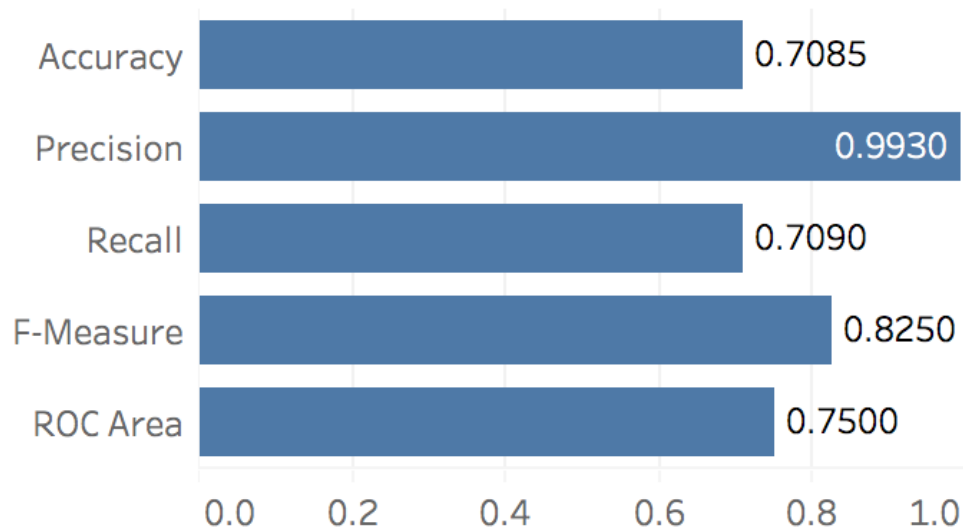
Confusion Matrix

0(no)	1(yes)	
8193	1869	0(no)
17	32	1(yes)

ROC Curve



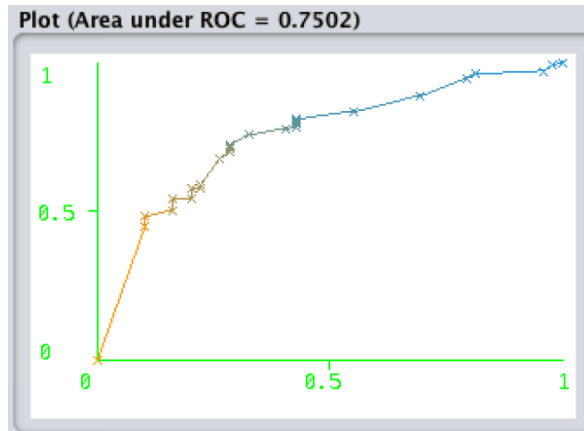
Random Tree



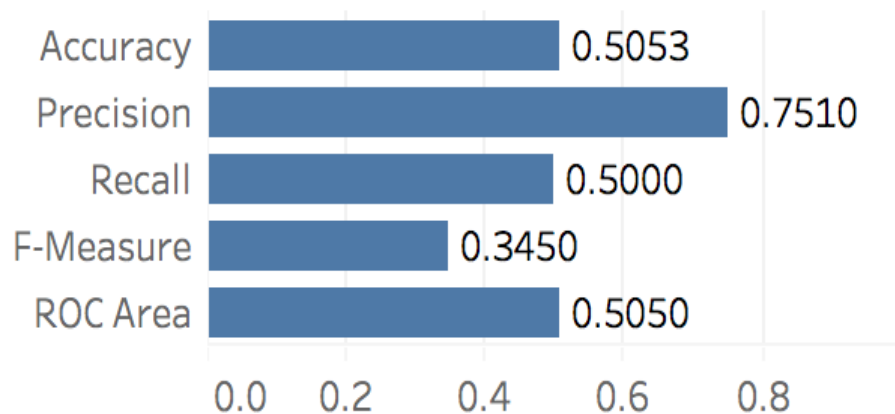
Confusion Matrix

0(no)	1(yes)	
7129	2933	0(no)
14	35	1(yes)

ROC Curve



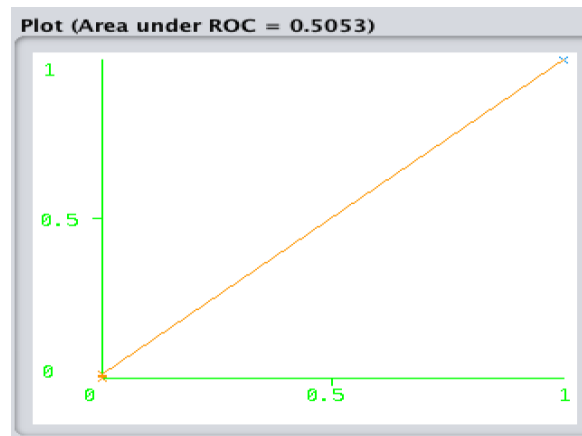
OneR



Confusion Matrix

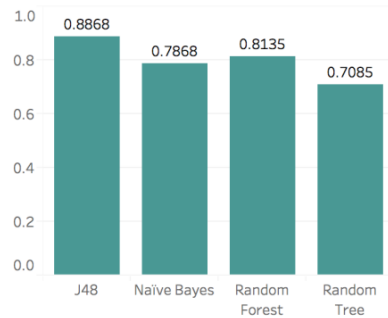
0(no)	1(yes)	
20219	2933	0(no)
20006	213	1(yes)

ROC Curve

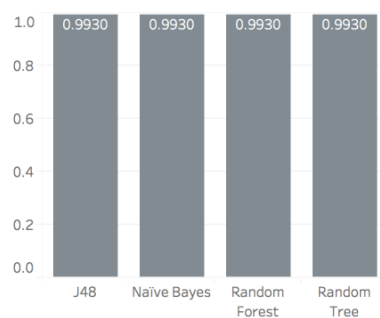


Performance Results

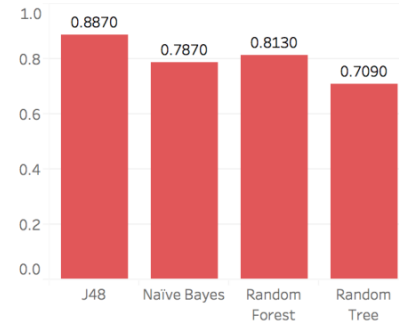
Accuracy



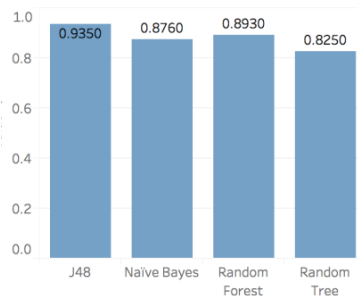
Precision



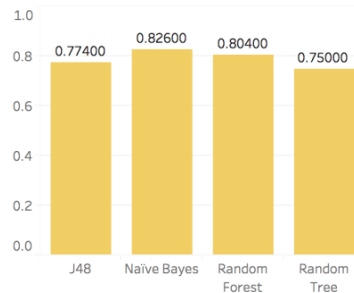
Recall



F-measure



Area under ROC curve



Results Discussion

Metrics

Accuracy
Precision
Recall

Best algorithm

C4.5 (J48)

Best attribute

Source

Conclusion

- Model is ready to use for predicting visitor engagement
- Many other bloggers or companies can use this same model with their attributes and classes for predicting the Visitor engagement with the blog.

Future Work

More attributes for more accurate prediction →
Efficient Model.

Some of the attributes but not limited -

- Search Keywords / Article Keywords
- Profession
- Interests
- Age etc.,

Thank You