

About Citation Policy Donate a Data Set Contact



Check out the beta version of the new UCI Machine Learning Repository we are currently testing! × Contact us if you have any issues, questions, or concerns. Click here to try out the new site.

# Facebook Comment Volume Dataset Data Set

Download: Data Folder, Data Set Description

**Abstract**: Instances in this dataset contain features extracted from facebook posts. The task associated with the data is to predict how many comments the post will receive.

Data Set Characteristics:	Multivariate	Number of Instances:	40949	Area:	N/A
Attribute Characteristics:	Integer, Real	Number of Attributes:	54	Date Donated	2016-03-11
Associated Tasks:	Regression	Missing Values?	N/A	Number of Web Hits:	113051

## Source:

Kamaljot Singh, Assistant Professor, Lovely Professional University, Jalandhar. Kamaljotsingh2009 '@' gmail.com

## **Data Set Information:**

The Dataset is uploaded in ZIP format. The dataset contains 5 variants of the dataset, for the details about the variants and detailed analysis read and cite the research paper

@INPROCEEDINGS{Sing1503:Comment, AUTHOR='Kamaljot Singh and Ranjeet Kaur Sandhu and Dinesh Kumar', TITLE='Comment Volume Prediction Using Neural Networks and Decision Trees', BOOKTITLE='IEEE UKSim-AMSS 17th International Conference on Computer Modelling and Simulation, UKSim2015 (UKSim2015)', ADDRESS='Cambridge, United Kingdom', DAYS=25, MONTH=mar, YEAR=2015, KEYWORDS='Neural Networks; RBF Network; Prediction; Facebook; Comments; Data Mining; REP Tree; M5P Trees.', ABSTRACT='The leading treads towards social networking services had drawn massive public attention from last one and half decade. The amount of data that is uploaded to these social networking services is increasing day by day. So, there is massive requirement to study the highly dynamic behavior of users towards these services. This is a preliminary work to model the user patterns and to study the effectiveness of machine learning predictive

modeling approaches on leading social networking service Facebook. We modeled the user comment patters, over the posts on Facebook Pages and predicted that how many comments a post is expected to receive in next H hrs. In order to automate the process, we developed a software prototype consisting of the crawler, Information extractor, information processor and knowledge discovery module. We used Neural Networks and Decision Trees, predictive modeling techniques on different dataset variants and evaluated them under Hits(at)10 (custom measure), Area Under Curve, Evaluation Time and Mean Absolute error evaluation metrics. We concluded that the Decision trees performed better than the Neural Networks under light of all evaluation metrics.'

}

The research paper is also available at conference website:

uksim.info/uksim2015/[Web Link]

another extended paper is that is to be published soon is :

@ARTICLE{Sing1601:Facebook,

AUTHOR='Kamaljot Singh',

TITLE='Facebook Comment Volume Prediction',

JOURNAL='International Journal of Simulation- Systems, Science and Technology-

IJSSST V16',

ADDRESS='Cambridge, United Kingdom',

DAYS=30,

MONTH=jan,

YEAR=2016,

KEYWORDS='Neural Networks; RBF Network; Prediction; Facebook; Comments; Data Mining; REP Tree; M5P Trees.',

ABSTRACT='The amount of data that is uploaded to social networking services is increasing day by day. So, their is massive requirement to study the highly dynamic behavior of users towards these services. This work is to model the user patterns and to study the effectiveness of machine learning predictive modeling approaches on leading social networking service Facebook. We modeled the user comment patters, over the posts on Facebook Pages and predicted that how many comments a post is expected to receive in next H hrs. To automate the process, we developed a software prototype consisting of the crawler, Information extractor, information processor and knowledge discovery module. We used Neural Networks and Decision Trees, predictive modeling techniques on different data-set variants and evaluated them under Hits(at)10, Area Under Curve, Evaluation Time and M.A.E metrics. We concluded that the Decision trees performed better than the Neural Networks under light of all metrics.'

}

this above paper will be freely available after publication at www.ijssst.info

## Attribute Information:

1 Page Popularity/likes Decimal Encoding Page feature Defines the popularity or support for the source of the document. 2 Page Checkins's Decimal Encoding Page feature Describes how many individuals so far visited this place. This feature is only associated with the places eg:some institution, place, theater etc.

#### 3

Page talking about Decimal Encoding Page feature Defines the daily interest of individuals towards source of the document/ Post. The people who actually come back to the page, after liking the page. This include activities such as comments, likes to a post, shares, etc by visitors to the page.

#### 4

Page Category Value Encoding Page feature Defines the category of the source of the document eg: place, institution, brand etc.

#### 5 - 29

Derived Decimal Encoding Derived feature These features are aggregated by page, by calculating min, max, average, median and standard deviation of essential features.

30 CC1 Decimal Encoding Essential feature The total number of comments before selected base date/time.

31 CC2 Decimal Encoding Essential feature The number of comments in last 24 hours, relative to base date/time.

32 CC3 Decimal Encoding Essential feature The number of comments in last 48 to last 24 hours relative to base date/time.

33 CC4 Decimal Encoding Essential feature The number of comments in the first 24 hours after the publication of post but before base date/time.

34 CC5 Decimal Encoding Essential feature The difference between CC2 and CC3.

35 Base time Decimal(0-71) Encoding Other feature Selected time in order to simulate the scenario.

36 Post length Decimal Encoding Other feature Character count in the post.

37 Post Share Count Decimal Encoding Other feature This features counts the no of shares of the post, that how many peoples had shared this post on to their timeline.

38 Post Promotion Status Binary Encoding Other feature To reach more people with posts in News Feed, individual promote their post and this features tells that whether the post is promoted(1) or not(0).

39 H Local Decimal(0-23) Encoding Other feature This describes the H hrs, for which we have the target variable/ comments received.

40-46 Post published weekday Binary Encoding Weekdays feature This represents the day(Sunday...Saturday) on which the post was published.

47-53 Base DateTime weekday Binary Encoding Weekdays feature This represents the day(Sunday...Saturday) on selected base Date/Time.

54 Target Variable Decimal Target The no of comments in next H hrs(H is given in Feature no 39).

### **Relevant Papers:**

Provide references to papers that have cited this data set in the past (if any). The Dataset is uploaded in ZIP format. The dataset contains 5 variants of the dataset, for the details about the variants and detailed analysis read and cite the

research paper

@INPROCEEDINGS{Sing1503:Comment,

AUTHOR='Kamaljot Singh and Ranjeet Kaur Sandhu and Dinesh Kumar'. TITLE='Comment Volume Prediction Using Neural Networks and Decision Trees', BOOKTITLE='IEEE UKSim-AMSS 17th International Conference on Computer Modelling and Simulation, UKSim2015 (UKSim2015)', ADDRESS='Cambridge, United Kingdom', DAYS=25, MONTH=mar, YEAR=2015. KEYWORDS='Neural Networks; RBF Network; Prediction; Facebook; Comments; Data Mining; REP Tree; M5P Trees.', ABSTRACT='The leading treads towards social networking services had drawn massive public attention from last one and half decade. The amount of data that is uploaded to these social networking services is increasing day by day. So, there is massive requirement to study the highly dynamic behavior of users towards these services. This is a preliminary work to model the user patterns and to study the effectiveness of machine learning predictive modeling approaches on leading social networking service Facebook. We modeled the user comment patters, over the posts on Facebook Pages and predicted that how many comments a post is expected to receive in next H hrs. In order to automate the process, we developed a software prototype consisting of the crawler, Information extractor, information processor and knowledge discovery module. We used Neural Networks and Decision Trees, predictive modeling techniques on different dataset variants and evaluated them under Hits(at)10 (custom measure), Area Under Curve, Evaluation Time and Mean Absolute error evaluation metrics. We concluded that the Decision trees performed better than the Neural Networks under light of all evaluation metrics.'

}

The research paper is also available at conference website:

uksim.info/uksim2015/[Web Link]

another extended paper is that is to be published soon is :

@ARTICLE{Sing1601:Facebook, AUTHOR='Kamaljot Singh', TITLE='Facebook Comment Volume Prediction', JOURNAL='International Journal of Simulation- Systems, Science and Technology-IJSSST V16'. ADDRESS='Cambridge, United Kingdom', DAYS=30, MONTH=jan, YEAR=2016, KEYWORDS='Neural Networks; RBF Network; Prediction; Facebook; Comments; Data Mining; REP Tree; M5P Trees.', ABSTRACT='The amount of data that is uploaded to social networking services is increasing day by day. So, their is massive requirement to study the highly dynamic behavior of users towards these services. This work is to model the user patterns and to study the effectiveness of machine learning predictive modeling approaches on leading social networking service Facebook. We modeled the user comment patters, over the posts on Facebook Pages and predicted that how many comments a post is expected to receive in next H hrs. To automate the process, we developed a software prototype consisting of the crawler, Information extractor, information processor and knowledge discovery module. We used Neural Networks and Decision Trees, predictive modeling techniques on different data-set variants and evaluated them under

Hits(at)10, Area Under Curve, Evaluation Time and M.A.E metrics. We concluded that the Decision trees performed better than the Neural Networks under light of all metrics.'

}

this above paper will be freely available after publication at www.ijssst.info

## **Citation Request:**

@INPROCEEDINGS{Sing1503:Comment, AUTHOR='Kamaljot Singh and Ranjeet Kaur Sandhu and Dinesh Kumar', TITLE='Comment Volume Prediction Using Neural Networks and Decision Trees'. BOOKTITLE='IEEE UKSim-AMSS 17th International Conference on Computer Modelling and Simulation, UKSim2015 (UKSim2015)', ADDRESS='Cambridge, United Kingdom', DAYS=25, MONTH=mar, YEAR=2015, KEYWORDS='Neural Networks; RBF Network; Prediction; Facebook; Comments; Data Mining; REP Tree; M5P Trees.', ABSTRACT='The leading treads towards social networking services had drawn massive public attention from last one and half decade. The amount of data that is uploaded to these social networking services is increasing day by day. So, there is massive requirement to study the highly dynamic behavior of users towards these services. This is a preliminary work to model the user patterns and to study the effectiveness of machine learning predictive modeling approaches on leading social networking service Facebook. We modeled the user comment patters, over the posts on Facebook Pages and predicted that how many comments a post is expected to receive in next H hrs. In order to automate the process, we developed a software prototype consisting of the crawler, Information extractor, information processor and knowledge discovery module. We used Neural Networks and Decision Trees, predictive modeling techniques on different dataset variants and evaluated them under Hits(at)10 (custom measure), Area Under Curve, Evaluation Time and Mean Absolute error evaluation metrics. We concluded that the Decision trees performed better than the Neural Networks under light of all evaluation metrics."

}

The research paper is also available at conference website:

uksim.info/uksim2015/[Web Link]

another extended paper is that is to be published soon is :

@ARTICLE{Sing1601:Facebook, AUTHOR='Kamaljot Singh', TITLE='Facebook Comment Volume Prediction', JOURNAL='International Journal of Simulation- Systems, Science and Technology-IJSSST V16', ADDRESS='Cambridge, United Kingdom', DAYS=30, MONTH=jan, YEAR=2016, UCI Machine Learning Repository: Facebook Comment Volume Dataset Data Set

KEYWORDS='Neural Networks; RBF Network; Prediction; Facebook; Comments; Data Mining; REP Tree; M5P Trees.',

ABSTRACT='The amount of data that is uploaded to social networking services is increasing day by day. So, their is massive requirement to study the highly dynamic behavior of users towards these services. This work is to model the user patterns and to study the effectiveness of machine learning predictive modeling approaches on leading social networking service Facebook. We modeled the user comment patters, over the posts on Facebook Pages and predicted that how many comments a post is expected to receive in next H hrs. To automate the process, we developed a software prototype consisting of the crawler, Information extractor, information processor and knowledge discovery module. We used Neural Networks and Decision Trees, predictive modeling techniques on different data-set variants and evaluated them under Hits(at)10, Area Under Curve, Evaluation Time and M.A.E metrics. We concluded that the Decision trees performed better than the Neural Networks under light of all metrics.'

}

this above paper will be freely available after publication at www.ijssst.info

Supported By:



In Collaboration With: Research + People -

About || Citation Policy || Donation Policy || Contact || CML