





The second component in this experiment was realized by applying several classifiers for bringing out the better classification accuracy for categorizing the subject catalog for the given dataset, namely “Topic Paper Author” dataset. It contains 18,375 instances and 5 attributes.

**Methods:**

The Matthews correlation coefficient (MCC) or phi coefficient is used in machine learning as a measure of the quality of binary (two-class) classifications. The proportion of correct predictions (also termed accuracy), are not useful when the two classes are of very different sizes. For example, assigning every object to the larger set achieves a high proportion of correct predictions, but is not generally a useful classification.

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

The above equation,

TP = True Positive,

TN = True Negative,

FP = False Positive and FN = False Negative.

Confusion Matrix

	Correct (A <sub>i</sub> )	Wrong (B <sub>j</sub> )
Correct (A <sub>i</sub> )	AB <sub>11</sub> [True Positive]	AB <sub>10</sub> [False Postive]
Wrong (B <sub>j</sub> )	AB <sub>01</sub> [False Negative]	AB <sub>00</sub> [True Negative]

The below machine learning classifications compute for finding the optimal classification algorithm for this research work.

- Bayes
- Lazy
- Meta
- Rules
- Trees

**III. RESULTS AND ANALYSIS**

In this section discusses results and analysis of this research work. This proposed work focuses on the computation and model optimization based on the one of the leading metrics namely Matthews correlation coefficient from the various machine learning algorithms like NaiveBayesMultinomialText classifier . It is under Bayes classifier. Then Instance based classifier or Lazy classifier. It belongs to Lazy category. Then AdaBoostM1 classifier belongs to Ensemble classifier , ZeroR classifier from Rules Based classifier and finally the DecisionStump classifier belongs to Trees classifier.

Table 3: Accuracy level for Various Classifiers

S.No	Category of the Classifier	Name of the Classifier	Accuracy
1	Bayes	NaiveBayeMultinomialText	64.59%
2	Lazy	IBK(k=1)	99.65%
3	Meta	AdaBoostM1	99.36%
4	Rules	ZeroR	64.60%
5	Trees	DecisionStump	72.22%

The above table represents the NaiveBayesMultinomialText classifier produces 64.59% level of accuracy, IBK classifier is 99.65% level of accuracy, AdaBoostM1 classifier is 99.36% and ZeroR classifier is 64.60% and DecisionStump classifier is 72.22%.

Table 4: Classifiers with Confusion Matrix Representation

S.No	Category of the Classifier	Name of the Classifier	Confusion Matrix
1	Bayes	NaiveBayeMultinomialText	$\begin{bmatrix} 11870 & 0 \\ 6505 & 0 \end{bmatrix}$
2	Lazy	IBK(k=1)	$\begin{bmatrix} 11860 & 10 \\ 54 & 6451 \end{bmatrix}$
3	Meta	AdaBoostM1	$\begin{bmatrix} 11869 & 1 \\ 300 & 6205 \end{bmatrix}$
4	Rules	ZeroR	$\begin{bmatrix} 11870 & 0 \\ 6505 & 0 \end{bmatrix}$
5	Trees	DecisionStump	$\begin{bmatrix} 11870 & 0 \\ 5106 & 1399 \end{bmatrix}$

The above table represents that the {TP,FP,FN,TN} for various algorithms. Namely, NaiveBayeMultinomialText classifier has {11870,0,6505,0}, IBK(K=1) classifier has {11860,10,54,6451}, AdaBoostM1 classifier has {11869,1,300,6205}, ZeroR Classifier has {11870,0,6505,0}, and DecisionStump Classifier has {11870,0,5106,1399}.

Table 5: Distribution of Matthews Correlation Coefficient

S.No	Category of the Classifier	Name of the Classifier	MCC(phi-coefficient ( $\phi$ ))
1	Bayes	NaiveBayeMultinomialText	0
2	Lazy	IBK(k=1)	0.99
3	Meta	AdaBoostM1	0.96
4	Rules	ZeroR	0
5	Trees	DecisionStump	0.39

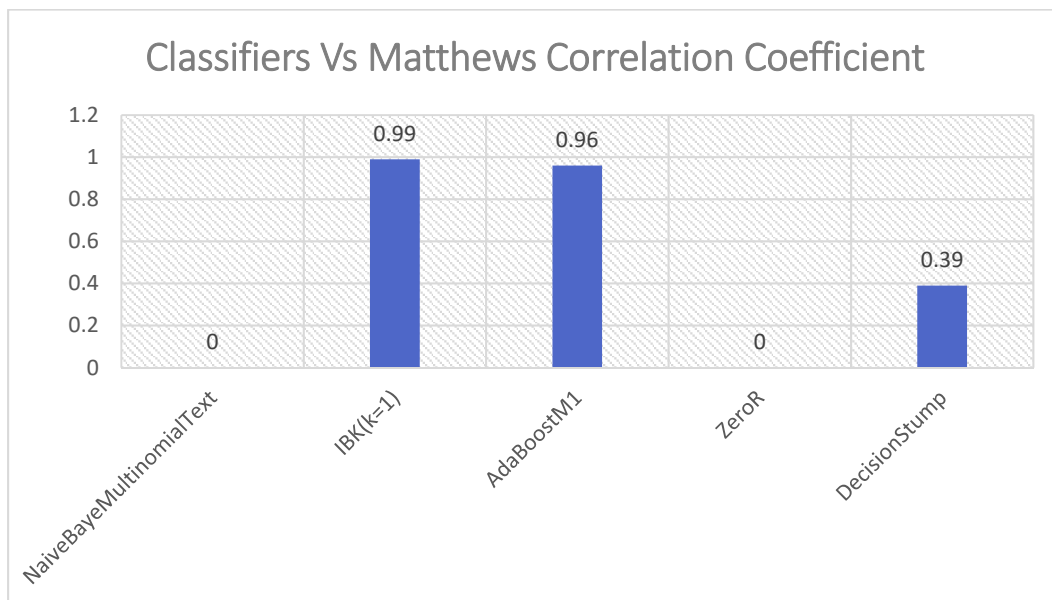


Figure 2 Graphical representations of various classifiers with their Correlation

The figure 2 clearly demonstrates that Matthews Correlation Coefficient values are various classifiers. Namely, the NaiveBayesMultinomialText classifier value is zero. The lazy classifier value is 0.99, The AdaBoostM1 classifier value is 0.96, ZeroR classifier value is Zero and DecisionStump Classifier value is 0.39.

So that it is representing the NaiveBayesMultinomialText belongs to Bayes Category classifier and ZeroR belongs to Rules category. These two classifiers are negatively correlated for this model. The DecisionStump is weakly correlated and Instance based classifier and AdaBoostM1 classifier are strongly correlated with positively.

#### IV. CONCLUSION

In this research work concludes that the based on the Matthews correlation coefficient metrics computed for confusion matrix of various leading machine learning algorithms, The DecisionStump algorithm, Instance based classifier and AdaBoostM1 Classifiers are correlated positively, but this proposed system recommends that AdaBoostM1 Classifier and IBK classifiers are strongly correlated with this model. And also Instance based classifier and AdaBoostM1 Classifiers are having above 99% accuracy level.

#### REFERENCES

- [1] Jie Tang, Sen Wu, Jimeng Sun, and Hang Su. Cross-domain Collaboration Recommendation. In Proceedings of the Eighteenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'2012).
- [2] Ayyappan.G et al. Knowledge Structure Infusion for Classification in Supervised Learning in Data Mining, International Journal on Computer Science and Engineering (IJCSE), Vol. 10 No.5 Apr-May 2018, Page No: 115-119,e-ISSN : 0975-3397, p-ISSN : 2229-5631.
- [3] <https://meshb.nlm.nih.gov/MeSHonDemand>
- [4] J. M. Hofman, A. Sharma, and D. J. Watts, "in Social Systems," vol. 488, no. February, pp. 486–488, 2017.
- [5] Ayyappan.G et al. Heart Disease Data Set Classifications: Comparisons of Correlation Co Efficient by Applying Various Parameters in Gaussian Processes, Indian Journal of Computer Science and Engineering (IJCSE), Volume No.9 Issue No.5 Oct-Nov 2018, Page No: 130-134, e-ISSN: 0976-5166, p-ISSN: 2231-3850.
- [6] <https://arxiv.org/pdf/0708.3601.pdf>
- [7] Ayyappan.G et al. Identification of Leading Research Contributors with Novel Performance Metrics Using Academic Social Network, International Journal on Computer Science and Engineering (IJCSE), Vol. 9 No.8 September 2017, Page No: 580-584,e-ISSN : 0975-3397, p-ISSN : 2229-5631.
- [8] <https://towardsdatascience.com/the-best-classification-metric-youve-never-heard-of-the-matthews-correlation-coefficient-3bf50a2f3e9a>
- [9] Ayyappan.G et al. A study on SNA: Text Mining using Academic Social Networks, International journal of Engineering and Technology (IJET), Volume No.8, Issue No.6, December 2016-Jan 2017, Page No: 2787-2790. ISSN: 0975-4024.
- [10] <http://arnetminer.org/lab-datasets/crossdomain/>
- [11] Ayyappan.G et al. A Novel K-NN Classification Approach Using Topic Modelling in Aminer Dataset, Indian Journal of Computer Science and Engineering (IJCSE), Volume No.10 Issue No.2 Apr-May 2019, Page No: 40-44, e-ISSN: 0976-5166, p-ISSN: 2231-3850.
- [12] <http://www.cs.waikato.ac.nz/ml/weka/>
- [13] Ayyappan.G et al. A Case Study on A Miner Dataset: Identifying leading research through various Models, Indian Journal of Computer Science and Engineering (IJCSE), Volume No.10 Issue No.3 Apr-May 2019, Page No: 45-53, e-ISSN: 0976-5166, p-ISSN: 2231-3850.
- [14] [https://en.wikipedia.org/wiki/Matthews\\_correlation\\_coefficient#Multiclass\\_case](https://en.wikipedia.org/wiki/Matthews_correlation_coefficient#Multiclass_case)
- [15] Ayyappan.G et al. Ensemble Classifications for Student Academics Performance Data Seta, Indian Journal of Computer Science and Engineering (IJCSE), Volume No.10 Issue No.1 Feb-Mar 2019, Page No: 31-34, e-ISSN: 0976-5166, p-ISSN: 2231-3850.