International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING

ISSN:2147-6799 www.ijisae.org Original Research Paper

Predicting Online Social Network Student Performance Using Enhanced RandomBayesian Algorithm

S. Senthamaraiselvi[1*], K. Meenakshi Sundaram[2]

Submitted: 17/09/2023 **Revised**: 18/11/2023 **Accepted**: 29/11/2023

Abstract: Online social networking sites, or OSNs, have become immensely popular in recent times. Many websites focus on locating and sharing various kinds of content as well as on making and keeping contacts. Statistical techniques for exploring data in educational settings and analyzing student performance are less effective than educational data mining. The research aims to employ various data mining approaches to investigate the effects of distinct variables on students' performance. This paper presents an Improved Mutual information based Filter Pearson's Correlation (IMIFPC) feature selection with Enhanced RandomBayesian (ERB) classification method for predicting the student performance. IMIFPC is a feature selection technique that uses combination of Filter method of Pearson's correlation with Mutual information method. The real-time OSN user dataset were used to test the proposed ERB approach to predict the students performance with classes (Excel and Vivekanandha) based on the data mining methods. The experimental results from the OSN User dataset demonstrate that the suggested approach using ERB classification performance achieved Accuracy of 98.00 and F1-Score of 97.99 percent.

Keywords: *Data mining, Classification, Random Forest, Feature Selection.*

1. Introduction

The task of extracting necessary data from a big database is known as data mining. It is the procedure for mining repositories for concealed information. In this sense, using machine learning to assist with data analysis and pattern recognition is a regulation. The model to be constructed is split into supervised and unsupervised learning based on the kind of data used and the learning challenge [1]. In unsupervised learning, there is simply input *x* and no output; in supervised learning, the training set consists of both input *x* and output *y*.

One strategy in that regard is the classification of predictive data mining. There, data is predicted based on certain features by utilizing a social media network data set for classification. In order to communicate with their social graph, other users, and the public, users can create, share, and participate in content on social media platforms. These platforms support a variety of content formats, including text, video, pictures, audio, PDFs, and PowerPoint. Social media can be defined as the channels of communication, services, and tools that help peers who share interests connect with one another [2].

Social media can be defined as the channels of communication, services, and tools that help peers who share interests connect with one other. Students may express their happiness [3], struggles, emotions, stress, and need for social support in a wonderful way on social media platforms like Face book, YouTube, Instagram, Twitter, and others. Students share and discuss their daily experiences informally and unstructured on a variety of social media platforms.

Educational researchers have historically gathered information about students' learning experiences through the use of techniques like reviews, surveys, interviews, focus groups, and classroom activities. These techniques can't be replicated or repeated frequently because they typically take a long time [4-5]. These investigations are typically small-scale as well. Furthermore, when asked about their experiences, students should consider what they were thinking and doing at the time, even though those details might need to be blurred over time. The process created to interpret social media data classification, as shown in figure 1.

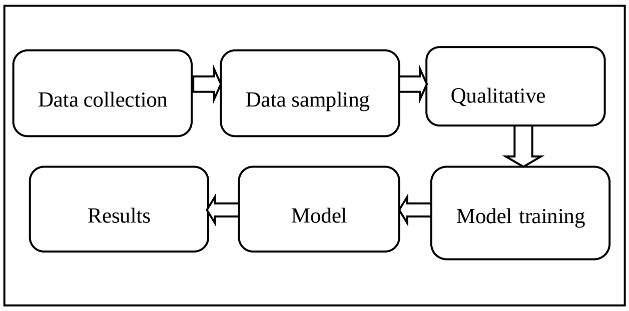


Fig. 1. The workflow of social media data.

[1] *Research Scholar (Part Time), Dept. of Computer Science, Erode Arts and Science College (Autonomous), Erode, Tamilnadu, India*
[2] *Associate Professor, Dept. of Computer Science, Erode Arts and Science College (Autonomous), Erode, Tamilnadu, India*
** Corresponding Author Email: sunruks_senthu@yahoo.co.in*

Web services known as Online Social Networks (OSN) enables users to share information with other network members and construct a public (or semi-public) profile containing some information about themselves [6]. The ability for users to explicitly establish links amongst them in the network is what makes them most distinctive. Social graphs, which are graphs that show these explicit linkages, are commonly used to evaluate OSN in a wide range of research.

One of the fundamental methods used in data mining procedures is classification. Assigning labels to the remaining dataset is the aim, given a set of labelled samples. Learning the desired properties of those samples can be achieved in this way. It is immediately apparent that the classification process can undermine user privacy, depending upon the nature of these inferred features.

2. Related Work

Boateng, R., and Amankwaa, A. [7] talked about the communication landscape of today's social environment is being rapidly altered by social media. Students' academic lives are being profoundly impacted by the emergence of social media. Academicians and institutions are constantly experimenting with social media technologies in an attempt to stimulate knowledge production, teamwork, and critical thinking abilities. These days, academic institutions embrace social media, allowing students to use it as a platform to interact with teachers, other students, and higher authorities worldwide. This necessitated an investigation of the effects of social media on students' academic lives. In evaluating these effects, the study pleaded for a qualitative method.

A novel student performance prediction model, known as the “student behavioral features,” was presented by Amrieh EA et al. [8] and was built on data mining techniques with new data attributes and features. These kinds of characteristics have to do with how learners interact with the e-learning platform. A combination of classifiers, including artificial neural networks, decision trees, and naïve bayesian models, assess the effectiveness of the student's predictive model. To enhance these classifiers' performance, the authors also used ensemble approaches. They employed the standard ensemble techniques seen in the literature: boosting, random forest (RF), and bagging. The results show that learners' behaviors and academic achievement are strongly correlated.

Using the Iterative Dichotomiser 3 (ID3), C4.5, and Classification and Regression tree (CART), Saheed YK et al. [9] proposed a method to predict student performance. Weka, or the Waikato Environment for Knowledge Analysis, was the site of the experiment. The ID3 accuracy, specificity, precision, recall, f-measure, and number of erroneously identified instances were all 95.9%, 93.8%, and 95.9%, respectively, according to the trial results. With respect to accuracy, specificity, precision, recall, f-measure, and number of erroneously identified instances, the C4.5 yielded results of 98.3%, 98.4%, 98.3%, and 1.70. According to the CART results, there were 1.70 wrongly classified instances, 98.3% accuracy, 98.3% specificity, 98.4% precision, 98.3% recall, and 98.3% f-measure.

According to Hussain et al. [10], educational data mining is essential for identifying students at an institution who struggle academically and for creating individualized recommendation systems for them. In their study, the authors took into account the records of students from three colleges in Assam, India, whose deep learning systems used the sequential neural model and the Adam optimization technique. To determine how well the authors could predict the students' findings, they examined alternative classification techniques including Adaboost and the Artificial Immune Recognition System v2.0. The deep learning methods produced the greatest categorization rate, 95.34%.

In a study conducted by Naicker et al. [11], the effectiveness of linear support vector machines was compared to the performance of the most advanced classical machine learning algorithms to see which method would perform better at predicting student performance. An experimental research design was employed in this quantitative investigation. Using a publicly accessible dataset of 1000 alpha-numeric student records, feature selection was used to build up the experiments. In predicting student achievement, linear support vector machines outperformed eleven categorical machine learning algorithms in a benchmarking study. According to the study's findings, lunch availability is the main factor influencing reading and writing performance, whereas characteristics like race, gender, and lunch have an impact on math achievement.

As demonstrated by Kumar et al. [12], Through the mining of educational data, the area of educational data mining seeks to uncover information and patterns in educational establishments. Anticipating the performance patterns of their students is essential for teachers to improve. A strategic strategy for providing high-quality education is one use for the knowledge acquired from it. Using data mining techniques and prior research, this article proposes that final student grades can be anticipated.

A thorough literature analysis on machine learning techniques for forecasting student performance and how to apply a prediction algorithm to determine the most significant attribute(s) in a student's data was presented by Stephen Opoku Oppong [13]. According to the study, neural networks are the most popular classifier for forecasting academic performance in pupils and also yield the most accurate findings. Additionally, 87% of the algorithms utilized were supervised learning algorithms, compared to

13% for unsupervised learning, and 59% of the research used different feature selection techniques to enhance the machine learning models' performance.

According to Jitender Kumar et al. [14], the amount of data in the educational system makes it harder to forecast students' success. Nowadays, the majority of educational institutions apply machine learning approaches to enhance their systems. These methods are used to analyze student performance and assist pupils in raising their level of performance. Thus, in order to comprehend machine learning techniques in education and learn how to forecast students' performance, a thorough examination of all relevant publications is required. The authors concentrated on identifying the critical elements that influence students' performance and discovered which prediction techniques are most frequently employed.

S. Senthamaraiselvi and K. Meenakshi Sundaram [15] discussed about the use of social media sites had affected performance of the students negatively and that there was a direct relationship between the use of social media sites abd academic performance. This study looked at the relationship between students' academic performance and the amount, kind, and usage of social networking sites in the classroom, as well as their exposure to these sites overall. A prediction framework that utilizes machine learning techniques, specifically Decision Tree (DT) and Random Forest (RF).

The research on the habits and motivations behind social network use among VIIMS students, conducted by S. Senthamaraiselvi and K. Meenakshi Sundaram [16], provides insight into the study's implementation and findings. A survey questionnaire that looked at user's habits and motivations for using Facebook, Instagram, Twotter, and Youtube was designed in order to perform the research. Using social media for educational purposes is one of the topics covered. After a rigorous analysis of the survey data collected over a period of 14 days, a comparison of the usage patterns and motivations for Facebook, Instagram, Twitter, and Youtube was made. Following a study of 140 survey responses, it was determined which social network college students most frequently used for both academic and recreational objectives.

3. Proposed Methodology

The proposed research methodology performs the OSN user data classification using data preprocessing, Improved Mutual information based Filter Pearson's Correlation (IMIFPC) and Enhanced RandomBayesian (ERB) Classification process is derived in this section. The OSN user data classification process considers overall process flow diagram is described in figure 2.

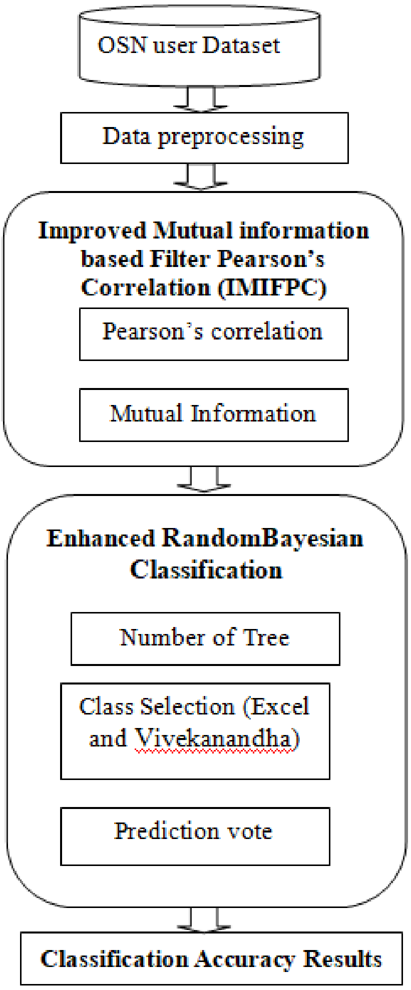
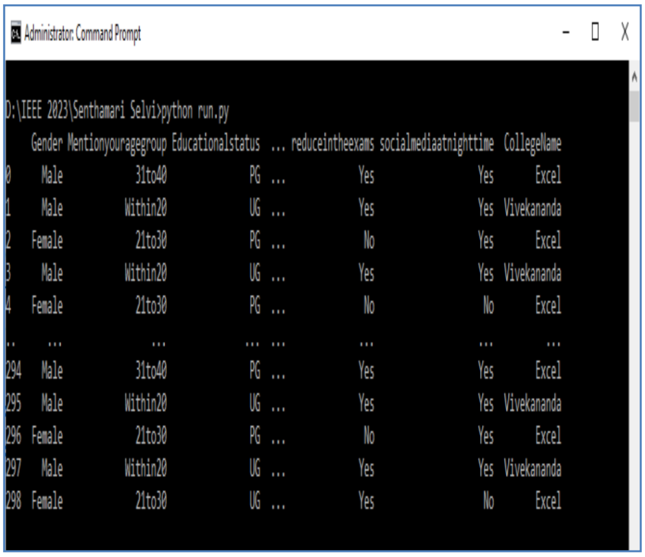

OSN user Dataset
Data preprocessing
Improved Mutual information based Filter Pearson's Correlation (IMIFPC)
Pearson's correlation
Mutual Information
Enhanced RandomBayesian Classification
Number of Tree
Class Selection (Excel and Vivekanandha)
Prediction vote
Classification Accuracy Results

Fig. 2. Proposed Flow Diagram

3.1. Dataset Preprocessing

One kind of data mining is data preprocessing, which involves turning unprocessed data into a comprehensible format. Since real-world data is frequently noisy, inconsistent, and incomplete, data preparation is crucial for both data warehousing and data mining. Cleaning, integrating, transforming, and reducing data are all part of data preparation.

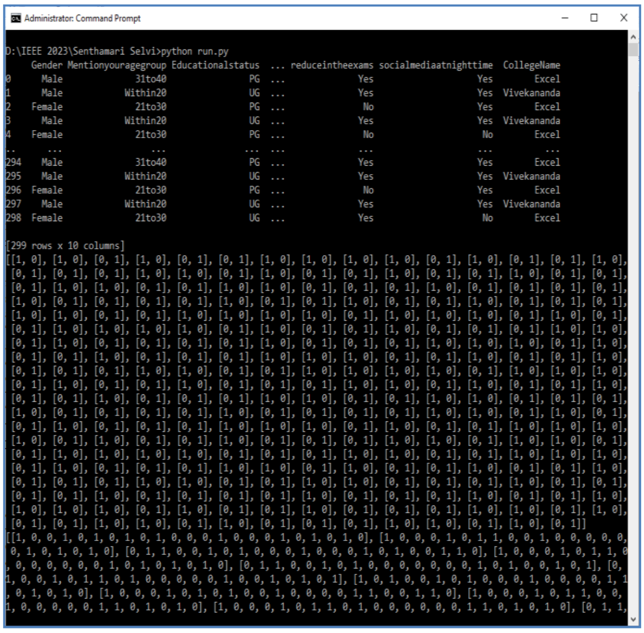
In the proposed OSN dataset have 10 features and 300 samples are contains incomplete, noisy and inconsistent are described in figure 3.



```
D:\IEEE 2023\Senthamari Selvi>python run.py
   Gender Mentionyouragegroup Educationalstatus  ... reduceintheexams socialmediaatnighttime  CollegeName
0    Male              31to40                PG  ...              Yes                   Yes        Excel
1    Male            Within20                UG  ...              Yes                   Yes  Vivekananda
2  Female              21to30                PG  ...               No                   Yes        Excel
3    Male            Within20                UG  ...              Yes                   Yes  Vivekananda
4  Female              21to30                PG  ...               No                    No        Excel
..    ...                 ...               ...  ...              ...                   ...          ...
294  Male              31to40                PG  ...              Yes                   Yes        Excel
295  Male            Within20                UG  ...              Yes                   Yes  Vivekananda
296 Female             21to30                PG  ...               No                   Yes        Excel
297  Male            Within20                UG  ...              Yes                   Yes  Vivekananda
298 Female             21to30                UG  ...              Yes                    No        Excel
```

Fig 3. OSN User Dataset

The complete features are transformed into attribute relation values following data preprocessing, as shown in figure 4 below.



```
D:\IEEE 2023\Senthamari Selvi>python run.py
   Gender Mentionyouragegroup Educationalstatus  ... reduceintheexams socialmediaatnighttime  CollegeName
0    Male              31to40                PG  ...              Yes                   Yes        Excel
1    Male            Within20                UG  ...              Yes                   Yes  Vivekananda
2  Female              21to30                PG  ...               No                   Yes        Excel
3    Male            Within20                UG  ...              Yes                   Yes  Vivekananda
4  Female              21to30                PG  ...               No                    No        Excel
..    ...                 ...               ...  ...              ...                   ...          ...
294  Male              31to40                PG  ...              Yes                   Yes        Excel
295  Male            Within20                UG  ...              Yes                   Yes  Vivekananda
296 Female             21to30                PG  ...               No                   Yes        Excel
297  Male            Within20                UG  ...              Yes                   Yes  Vivekananda
298 Female             21to30                UG  ...              Yes                    No        Excel

[299 rows x 10 columns]
[[1, 0], [1, 0], [0, 1], [1, 0], [0, 1], [0, 1], [1, 0], [1, 0], [1, 0], [1, 0], [0, 1], [1, 0], [0, 1], [0, 1], [1, 0],
 [0, 1], [0, 1], [0, 1], [1, 0], [1, 0], [0, 1], [1, 0], [0, 1], [0, 1], [1, 0], [0, 1], [0, 1], [1, 0], [0, 1], [0, 1],
 [0, 1], [1, 0], [1, 0], [0, 1], [0, 1], [1, 0], [1, 0], [0, 1], [0, 1], [0, 1], [1, 0], [0, 1], [0, 1], [0, 1], [0, 1],
 [1, 0], [0, 1], [0, 1], [0, 1], [0, 1], [1, 0], [0, 1], [0, 1], [1, 0], [0, 1], [0, 1], [1, 0], [0, 1], [0, 1], [0, 1],
 [1, 0], [1, 0], [0, 1], [0, 1], [1, 0], [1, 0], [1, 0], [1, 0], [0, 1], [1, 0], [1, 0], [1, 0], [0, 1], [1, 0], [0, 1],
 [0, 1], [1, 0], [1, 0], [1, 0], [1, 0], [0, 1], [1, 0], [0, 1], [0, 1], [1, 0], [0, 1], [0, 1], [0, 1], [1, 0], [1, 0],
 [0, 1], [1, 0], [0, 1], [0, 1], [0, 1], [0, 1], [1, 0], [0, 1], [0, 1], [0, 1], [1, 0], [1, 0], [1, 0], [0, 1], [1, 0],
 [0, 1], [0, 1], [1, 0], [0, 1], [0, 1], [0, 1], [1, 0], [1, 0], [1, 0], [0, 1], [0, 1], [1, 0], [0, 1], [0, 1], [1, 0],
 [0, 1], [0, 1], [0, 1], [1, 0], [1, 0], [0, 1], [0, 1], [1, 0], [1, 0], [1, 0], [1, 0], [0, 1], [1, 0], [1, 0], [1, 0],
 [0, 1], [1, 0], [0, 1], [0, 1], [1, 0], [0, 1], [0, 1], [0, 1], [1, 0], [1, 0], [0, 1], [1, 0], [0, 1], [0, 1], [1, 0],
 [0, 1], [0, 1], [1, 0], [0, 1], [0, 1], [0, 1], [1, 0], [1, 0], [0, 1], [0, 1], [1, 0], [1, 0], [0, 1], [0, 1], [0, 1],
 [1, 0], [0, 1], [0, 1], [0, 1], [0, 1], [1, 0], [0, 1], [0, 1], [0, 1], [0, 1], [1, 0], [1, 0], [0, 1], [1, 0], [0, 1],
 [0, 1], [1, 0], [0, 1], [0, 1], [0, 1], [1, 0], [1, 0], [1, 0], [0, 1], [1, 0], [0, 1], [0, 1], [1, 0], [1, 0], [1, 0],
 [1, 0], [0, 1], [1, 0], [0, 1], [0, 1], [1, 0], [0, 1], [0, 1], [0, 1], [1, 0], [1, 0], [0, 1], [1, 0], [1, 0], [0, 1],
 [0, 1], [1, 0], [1, 0], [1, 0], [1, 0], [0, 1], [0, 1], [0, 1], [1, 0], [1, 0], [0, 1], [0, 1], [1, 0], [1, 0], [1, 0],
 [0, 1], [1, 0], [0, 1], [0, 1], [0, 1], [1, 0], [1, 0], [0, 1], [1, 0], [0, 1], [0, 1], [1, 0], [0, 1], [0, 1], [1, 0],
 [0, 1], [0, 1], [0, 1], [1, 0], [1, 0], [0, 1], [0, 1], [1, 0], [1, 0], [0, 1], [0, 1], [0, 1], [1, 0], [0, 1], [0, 1],
 [0, 1], [0, 1], [1, 0], [0, 1], [0, 1], [0, 1], [0, 1], [1, 0], [0, 1], [0, 1], [0, 1], [1, 0], [0, 1], [0, 1], [1, 0],
 [1, 0], [1, 0], [1, 0], [0, 1], [1, 0], [0, 1], [0, 1], [1, 0], [0, 1], [0, 1], [0, 1], [1, 0], [1, 0], [0, 1], [1, 0],
 [0, 1], [0, 1], [1, 0], [1, 0], [0, 1], [1, 0], [0, 1], [0, 1], [0, 1], [1, 0], [1, 0], [0, 1], [1, 0], [0, 1]]
[[1, 0, 0, 1, 0, 1, 0, 1, 0, 1, 0, 0, 0, 1, 0, 0, 0, 1, 0, 1, 0, 1, 0], [1, 0, 0, 0, 1, 0, 1, 1, 0, 0, 1, 0, 0, 0, 0, 0,
 0, 1, 0, 1, 0, 1, 0], [0, 1, 1, 0, 0, 1, 0, 1, 0, 0, 0, 1, 0, 0, 0, 1, 0, 1, 0, 0, 1, 1, 0], [1, 0, 0, 0, 1, 0, 1, 1, 0
, 0, 0, 0, 0, 0, 0, 1, 0, 1, 0, 1, 0, 1, 0], [0, 1, 1, 0, 0, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0, 1, 0, 1, 0, 0, 1, 0, 1], [0,
1, 0, 0, 1, 0, 1, 1, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 1, 0, 1, 0, 1], [1, 0, 1, 0, 0, 1, 0, 1, 0, 0, 0, 1, 0, 0, 0, 0, 1, 1
, 0, 1, 0, 1, 0], [1, 0, 0, 0, 1, 0, 1, 0, 1, 0, 0, 1, 0, 0, 0, 0, 1, 1, 0, 0, 1, 1, 0], [1, 0, 0, 0, 1, 0, 1, 1, 0, 0,
1, 0, 0, 0, 0, 0, 1, 1, 0, 1, 0, 1, 0], [1, 0, 0, 0, 1, 0, 1, 1, 0, 1, 0, 0, 0, 0, 0, 0, 1, 1, 0, 1, 0, 1, 0], [0, 1, 1,
```

Fig 4. OSN User Preprocessed data

3.2. Improved Mutual information based Filter Pearson's Correlation (IMIFPC)

Finding the best feature within a data collection is the aim of feature selection. Data can be categorized by machine learning algorithms into a group of class attributes and goals. Feature domains containing hundreds of variables or features are now used in machine learning applications. The problem of removing pointless and irrelevant variables has been tackled in a number of ways. Feature selection (attribute or variable reduction) enhances performance, lowers processing expenses, lessens the effects of the "dimensional curse," and improves data comprehension.

The proposed Improved Mutual Information based Filter Pearson's Correlation (IMIFPC) method looks for examining the properties of each individual feature and utilizing MIFPC as a selection criterion. It is a metric for measuring the degree of linear dependence between X and Y, two continuous variables. It has a value between -1 and +1. The value of Pearson's correlation is:

$$PC_{X,Y} = \frac{cov(X,Y)}{\sigma_X \sigma_Y} \quad (1)$$

The overall selection procedure is detailed in Algorithm 1. It is a heuristic incremental search approach in which the selection procedure is repeated until the desired number of b input characteristics is chosen. To remove the burden of choosing an acceptable value for the redundancy parameter, propose a new feature selection criterion, and find a feature relevance FR in equation (2). It concurrently maximizes FC (*OF*, *feat*$_i$) and minimizes average redundancy with perarson's correlation PC's by selecting a feature from a given input feature collection.

$$FR = FC(OF, feat_i) - \frac{1}{|S|} \sum_{f_{s \in Sel}} PC \quad (2)$$

Whereas, *FR* represents Feature relevance; *FC* represents Feature correlation; *OF* represents Operational Feature; *feat*$_i$ represents feature; S signifies Selection; *PC* represents Pearson's correlation. Figure 5 illustrates IMIFPC feature result.

Algorithm 1: Improved Mutual information based Filter Pearson's Correlation (IMIFPC)

Input: Feature subset *Fsb*= {*fsb*$_1$, *fsb*$_2$, …,*fsb*$_n$}, *nf*-Number of features

Output: S_f ←Selected features

Process

```
1: Initialization: S = 0 ;
2: initiate FC(OF, feati) for every feature, i = 1, 2, …,n
3: choose feature $feat_i$ that maximise,
   $\underset{feat_i}{argmax} \left(FC(OF, feat_i)\right), i = 1, 2, \cdots, n_f$    (3)
   $FSB \leftarrow FSB\{feat_i\}$    (4)
   $S \leftarrow \{feat_i\}$    (5)
4: While (|S|<nf ) do
```

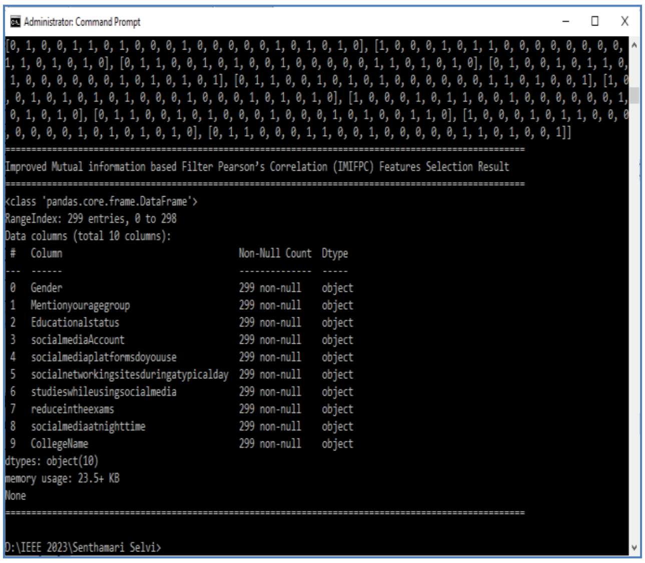
$$FR = \underset{feat_i \in FSB}{argmax} \left(FC(OF, feat_i)\right) - PC \sum_{f_s \epsilon S} IF(f_{eati}, f_s) \quad (6)$$

$$FSB \leftarrow FSB\{f_i\} \quad (7)$$

$$S \leftarrow \{feat_i\} \quad (8)$$

end

return S_f



```
[0, 1, 0, 0, 1, 1, 0, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0, 1, 0, 1, 0, 1, 0], [1, 0, 0, 0, 1, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0,
1, 1, 0, 1, 0, 1, 0], [0, 1, 1, 0, 0, 1, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 1, 1, 0, 1, 0, 1, 0], [0, 1, 0, 0, 1, 0, 1, 1, 0,
 1, 0, 0, 0, 0, 0, 0, 1, 0, 1, 0, 1, 0, 1], [0, 1, 1, 0, 0, 1, 0, 1, 0, 1, 0, 0, 0, 0, 0, 0, 1, 1, 0, 1, 0, 0, 1], [1, 0
, 0, 1, 0, 1, 0, 1, 0, 1, 0, 0, 0, 1, 0, 0, 0, 1, 0, 1, 0, 1, 0], [1, 0, 0, 0, 1, 0, 1, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 1,
 0, 1, 0, 1, 0], [0, 1, 1, 0, 0, 1, 0, 1, 0, 0, 0, 1, 0, 0, 0, 1, 0, 1, 0, 0, 1, 1, 0], [1, 0, 0, 0, 1, 0, 1, 1, 0, 0, 0
, 0, 0, 0, 0, 1, 0, 1, 0, 1, 0, 1, 0], [0, 1, 1, 0, 0, 0, 1, 1, 0, 0, 1, 0, 0, 0, 0, 0, 1, 1, 0, 1, 0, 0, 1]]
==========================================================================================
Improved Mutual information based Filter Pearson's Correlation (IMIFPC) Features Selection Result
==========================================================================================
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 299 entries, 0 to 298
Data columns (total 10 columns):
 #   Column                                Non-Null Count  Dtype
---  ------                                --------------  -----
 0   Gender                                299 non-null    object
 1   Mentionyouragegroup                   299 non-null    object
 2   Educationalstatus                     299 non-null    object
 3   socialmediaAccount                    299 non-null    object
 4   socialmediaplatformsdoyouuse          299 non-null    object
 5   socialnetworkingsitesduringatypicalday 299 non-null   object
 6   studieswhileusingsocialmedia          299 non-null    object
 7   reduceintheexams                      299 non-null    object
 8   socialmediaatnighttime                299 non-null    object
 9   CollegeName                           299 non-null    object
dtypes: object(10)
memory usage: 23.5+ KB
None
==========================================================================================


D:\IEEE 2023\Senthamari Selvi>
```

Fig 5. IMIFPC Features Result

3.3. Enhanced RandomBayesian (ERB) Classification

This research presents a novel method of Enhanced RandomBayesian classification is an extension of Random Forest and Bayesian Classification method. The first step in this classification is to initialize the size and number of tree variables. The entire preprocessed feature set is used as the training feature, while the final feature selection is used as the test feature. To acquire the feature score of feature dimension, this technique preprocesses the training and testing features. Several FOLD (First Order Logical Decision) methods are used to classify an ERB classification case in order to assess the prediction class. With its ten features, the random selection feature categorization reaches the class (Excel and Vivekanandha) colleges.

$$Sel_{feat} = \sum_{k=1}^{class} \sum_{j=1}^{feat} find\left(Vote_{feat} = prediction_{class}\right) \quad (9)$$

Where *feat* is features; *class* (*Excel* and *Vivekanandha*); Vote is unique features in the trained dataset.

In order to acquire a new set of features, one must first discretize the feature values by finding the relevant feature and then apply Equation (10), the classification formula, to calculate the continuous values.

$$Class_k = \frac{1}{\sqrt{2\pi\sigma_k^2}} \quad (10)$$

where σ^2_k be the variance of features with class $Class_k$ (Excel and Vivekanandha).

Following selection, the ERB procedure achieves classification accuracy below equation 11.

$$Classification_{acc} = \sum_{m=1}^{len(vote)} Sel(Class_m) \quad (11)$$

4. Experimental Results

The proposed Enhanced RandomBayesian (ERB) method was used to estimate the results. The results are accomplished using Intel I5-6500U series 3.28GHz x64-based processors, 8GB main memory, and python 3.8 simulations on the Windows 10 operating system. The Precision, recall, accuracy and F1-score for each model result of ERB Classification. Table 1 describes the comparison of precision, recall, accuracy and F1-score measures with conventional methods.

$$Precison = \frac{TP}{TP + FP} \quad (12)$$

$$Recall = \frac{TP}{TP + FN} \quad (13)$$

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (14)$$

$$F1score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (15)$$

Table 1. Comparison of precision, recall, accuracy and F1-score measures.

Methods	Precision	Recall	Accuracy	F1-Sore
DT	87.5	100	90.00	89.33
Random Forest	92.85	100	95.00	94.90

Naïve Bayes (NB)	97.28	98.69	96.84	96.84
Proposed ERB	**96.77**	**100**	**98.00**	**97.99**

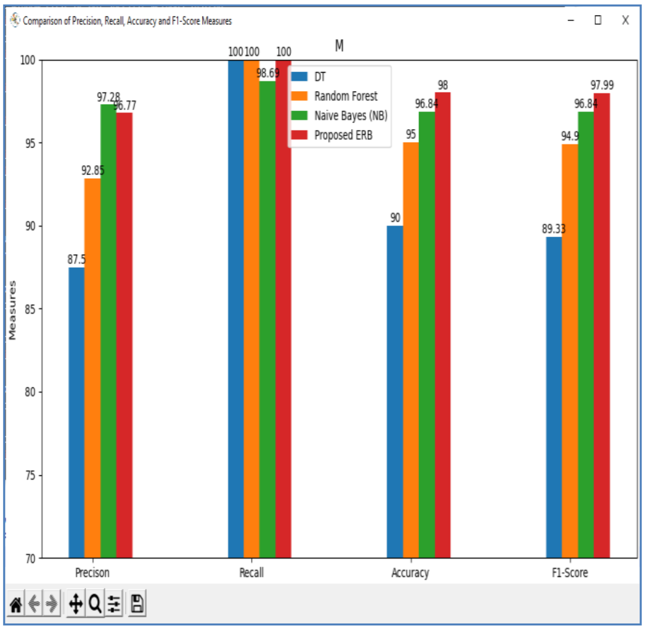


Fig 6. Performance Analysis

Figure 6 shows that the suggested approach verified that the Accuracy and F1-score rise occurred at a maximum value of 98 and 97.99 percent.

5. Conclusion and Future Work

The paper proposed Improved Mutual information based Filter Pearson's Correlation (IMIFPC) feature selection with Enhanced RandomBayesian classification method in this study, and tested it on an OSN real-world datasets. The proposed ERB Classification method provides superior results compared with conventional algorithms in speed and accuracy terms. In this research, three procedures were used in its design and development: (1) data preprocessing; (2) IMIFPC feature selection; and (3) ERB classification. The first step involves preprocessing the data to convert the trained statistical data. The second step involves choosing the best feature selection method to predict the social network data. The last step predicts the ERB (RP) classification to estimate accuracy. Thus, it can be inferred that the suggested approach lowers computational expenses and presents a compelling substitute for reducing the dimensionality of data while maintaining a high degree of accuracy. As a result, effective classification throughout the execution of a process is required, which is a critical concern in future work. Examining the effectiveness of deep learning algorithms is another possible avenue for future research.

References

[1] W. W. Yaacob, N. M. Sobri, S. M. Nasir, N. D. Norshahidi, and W. W. Husin, "Predicting student drop-out in higher institution using data mining techniques," Journal of Physics: Conference Series, vol. 1496, p. 012005, 2020.

[2] P. Sokkheyand and T. Okazaki, "Developing web-based support systems for predicting poor-performing students using educational data mining techniques," Studies, vol. 11, no. 7, 2020.

[3] Xin Chen, Mihaela and Krishna P.C, "Mining Social Media Data for Understanding Students Learning Experiences", IEEE Transactions on Learning Technologies, 2014.

[4] X. Xu, J. Wang, H. Peng, and R. Wu, "Prediction of academic performance associated with internet usage behaviors using machine learning algorithms," Computers in Human Behavior, vol. 98, pp. 166–173, 2019.

[5] A. Alhassan, B. Zafar, and A. Mueen, "Predict students' academic performance based on their assessment grades and online activity data," International Journal of Advanced Computer Science and Applications (IJACSA), vol. 11, no. 4, 2020

[6] Siemens and P. Long, "Penetrating the fog: Analytics in learning and education," Educause Review, vol. 46, no. 5, pp. 30–32, 2011.

[7] Boateng, R., & Amankwaa, A. (2016). The impact of social media on student academic life in higher education. Global Journal of Human-Social Science, 16(4), 1-8.

[8] Amrieh EA, Hamtini T, Aljarah I. "Mining educational data to predict student's academic performance using ensemble methods." International Journal of Database Theory and Application. 2016;9(8):119–136.

[9] Saheed YK, Oladele TO, Akanni AO, Ibrahim WM. "Student performance prediction based on data mining classification techniques." Nigerian Journal of Technology. 2018;37(4):1087.

[10] Hussain S, Muhsion ZF, Salal YK, Theodoru P, Kurtouglu F, Hazarika GC. "Prediction model on student performance based on internal assessment using deep learning." International Journal of Emerging Technologies in Learning ({IJET}). 2019;14(08):4.

[11] Naicker N, Adeliyi T, Wing J. "Linear support vector machines for prediction of student performance in school-based education." Math Probl Eng. 2020

[12] Kumar M, Sharma C, Sharma S, Nidhi N, Islam N.

"Analysis of feature selection and data mining techniques to predict student academic performance." In 2022 International Conference on Decision Aid Sciences and Applications (DASA), IEEE. 2022:1013–1017.

[13] Stephen Opoku Oppong, "Predicting Students' Performance Using Machine Learning Algorithms: A Review", Asian Journal of Research in Computer Science, Volume 16, Issue 3, 2023.

[14] Jitender Kumar, Ritu Vashistha, Kushwant Kaur and Siroj Kumar Singh, "Machine Learning Techniques of Predicting Student's Performance", International Conference in Advances in Power, Signal, and Information Technology (APSIT), IEEE, 2023.

[15] S. Senthamaraiselvi and K. Meenakshi Sundaram, "Reprecussion of Social Media in the Area of Education", Inernational Conference on Advanced Computing (ICAC), 2023.

[16] S. Senthamaraiselvi and K. Meenakshi Sundaram, "A study on social networking usage amoung students", 4th International Conference on Artificial Intelligence Trending Towards Automation, 2023.

Mrs. S.Senthamaraiselvi Pursuing Ph.D Computer science, Part time Research scholar in Erode Arts and Science College and is presently, Assistant professor in the Department of MCA at Vivekanandha Institute of Information and Management Studies, Tiruchengode. She Has 18 years of academic experience. she has published 5 papers in International Journals and has more than 20 presentaions in International and National conferences.

Dr.K.Meenakshi Sundaram is presently, Associate Professor and Head, PG and Research Dept. of Computer Science, Erode Arts and Science College (Autonomous), Erode. He received his Ph.D.(part timre) in Computer Science from D.G.Vaishnav College, Chennai, He holds his Master of Computer Applications (MCA) from PSG College of Tech, Coimbatore and Master of Science (Mathematics) from PSG College of Arts and Science, Coimbatore. He has published papers in 37 International journals (Scopus Indexed and UGC Approved) and 35 National Journals and has 55 presentations in International and National conferences to his credit. Has 33 years of academic experience and has held various positions at Erode Arts and Science College (Autonomous), Erode.