

Ambo University Student's Case Classification Models using Support Vector Machine

Naol Bakala Defersha

Abstract: *The main objective of ambo university is to provide quality education and improve the overall performance of an students by looking at individual students' problems cases. One way to analysis students' cases personally is to identify the problems causes and guide the students to solve the problems. Following this, the department Academic council and Academic Commission is whole authorized people to make the decision manually so this will consume more time and energy. This research focused to learning classification models for predicting students problems cases using support vector classification techniques. Finally, performance of the model evaluated using precision, recall and F-measure evaluation parameters.*

Keywords: *Decision making, Evaluation Parameters, Machine learning algorithms, Prediction Model, student cases, Support Vector approach.*

I. INTRODUCTION

Ambo University is one of government University found in Ambo town, Oromia, Ethiopia and It located on West direction on 111KM from Addis Ababa [1]. It was opened as University in May 2011 by Ethiopian Government. Today, Ambo University is hosting Undergraduate and Graduate Programs in Extension and Regular mode. Students in Ambo University faced with different challenges that have to be solved by University concerned body like Department Academic council, Academic Commission (AC) and even University Senate at Large depending on the problem of students. To get immediate help/service students applied applications to concerned body in Ambo University. Law highly connected with life of society in any organization [2]. Concerned body in University has to discuss on the case of students as rule and regulation of University and Ministry of Science and Higher Education laws. In Ambo University there is principle by which students case reviewed by department Academic Council (DAC), and Academic Commission (AC) to give decision to student case. To give right decision on students' cases each committee considered and used working principle of University. even if, the decision maker used already set rule and regulation to give accurate, efficient and high-speed decision they faced various challenges such as inconsistent decision, resources consumption and tedious. those challenges raised due to the decision making is manual ways.

As a solution, it is essential to apply advantages of machine Learning approach to build student case Prediction model to automate student's case decision making tasks. As indicate in work of [3].

A support vector machine is machine learning algorithm used to design prediction model and score good result. To accomplish building the Prediction model researcher aimed to prepare student case perdition data set, design Architecture of model that predict result of student's case application, to design model that predict student case result depend on provided case and evaluate performance of model and forward for future work. As described in section (II) data manually collected and prepared by researcher from Ambo University Registrar, Departments, Legislation and Policy/laws of Ministry of Science and Higher Education. Data collected were in only Text document and not Include other type video, audio, image data and etc. The Experiment of conducted by loading data set into python programming Language. Researcher collected on Text data that excluded data informs of video, audio or image. After experimented conducted and model generated the accuracy of model displayed in precision, recall and F-measure evaluation parameters. In this experiment the F-measure/f-score resulted with 98 % average Value.

II. METHODOLOGY

Methodology is series of techniques used by researcher to achieve research objectives [4]. Some researches methodologies utilized in particular research works are literature review. The literature review used in this work include some Wikipedia and rule and Regulation of University whereas most of references used were collected from journal Articles relevant to this work. Data collection and Data set Preparation techniques also used as main part of methodology in research. As indicated in section... data were collected from Ambo University Registrar office, Ministry of Higher education of Ethiopia and Ambo University Legislation. Data collected were in word format. After data collected, its different tasks performed on it and data set prepared to conduct experiment. After experiment conducted, it is essential to evaluate the performance of built model using evaluation Techniques. As indicated in section (II) precision, recall and f-measure parameters were used to evaluate model. Student case prediction Model was developed to check the output of this research.

A. Data collection and Data set Preparation

There is no standard data set prepared to predict students' case in Ambo University. Due to this, researcher manually prepared data set by collecting data from Ambo University Registrar records, Ministry of Science and Higher education of Ethiopia and Ambo University Legislation.

Revised Manuscript Received on June 22, 2020.

Naol Bakala Defersha, Lecturer and Administrative Vice Director, Ambo University, Institute of Technology.

Data collected in word format. Researcher does not collect video, image, audio and graphics information about students. After data collected, data converted to CSV format that suitable to apply machine learning approach. All data collected are not equally important, some features are important and others are not important. To obtain clean data from unimportant data from collected data researcher applied feature extraction and feature Selection technique. Author of [2] used Vectorization approach to extract feature. In this research Vectorization techniques used to discover important feature. Before loading our data set into Python Programming language program for experiment identify the three main label as class. Accepted, Pending and Rejected case are three main labels nominated as a class in data set. Accepted case replaced by "0", pending case replaced by "1" and Rejected case replaced by "2". After data set preparation finalized, data set loaded into python for experiment. Prepared data set also split into training and test set. As elaborated by [2] data set split into 80% or 70% training set and 20% or 30% test set. Increasing volume training data set increase accuracy, the accuracy of the prediction Model, in this work data set divided into train set (80%) and test set (20%) to train model and test its performance. Researcher used Python Language to build student case result prediction model.

Evaluation

Evaluating the performance of built model is one step in this research. In this work the performance of student case prediction model using evaluation precision, recall and f-measure are evaluation parameters used to evaluate model. F-measure is the most accurate since

F-measure/score is average of it is average of precision and recall [5]. As it indicated in [5] it represented by

$$F_1 = \frac{2pr}{p+r}$$

. Since it is the average of precision and recall researcher used F1-score to examine the accuracy of built model.

Support Vector Machine (SVM) is "linear or nonlinear classifier" that used to classify two objects [6]. It is well known supervised machine learning algorithm that widely used to build classifier model. In this research after data set divided into train and test set, SVM algorithm applied on train set to train model.

	precision	recall	f1-score	support
0	1.00	0.98	0.99	104
1	1.00	1.00	1.00	1
2	0.93	1.00	0.96	26
accuracy			0.98	131
macro avg	0.98	0.99	0.98	131
weighted avg	0.99	0.98	0.98	131

Figure 2 student case Prediction model result Using SVM

III. EXPERIMENTAL RESULT

The implementation of student case prediction using Support Vector machine in Python Language.

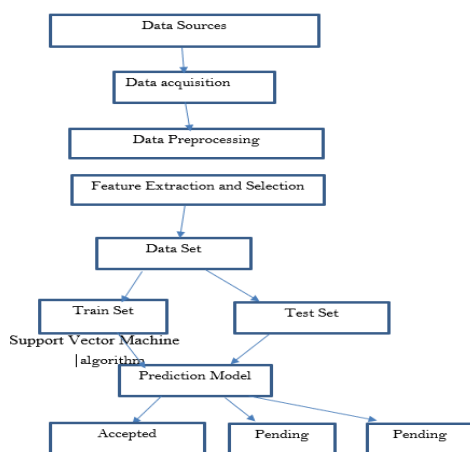


Figure 1 Student case Result Prediction Model Designing Architecture Support Vector Machine

```
In [196]: %matplotlib inline
import pandas as pd
import matplotlib.pyplot as plt
```

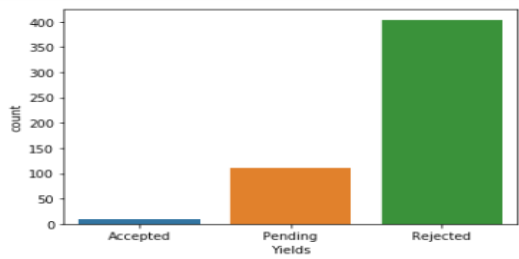
```
In [197]: ST = pd.read_csv('C:/Users/Mine/Desktop/Publication - 1/Naol publication - 1/ST-D-20-F-CSV.csv')
ST.head()
```

Out[197]:

```
In [199]: print(ST['class'].unique())
[1 2 0]
```

```
In [200]: print(ST.groupby('Yields').size())
Yields
Accepted    10
Pending    110
Rejected    404
dtype: int64
```

```
In [201]: import seaborn as sns
sns.countplot(ST['Yields'], label="Count")
plt.show()
```



```
In [202]: ST.describe()
```

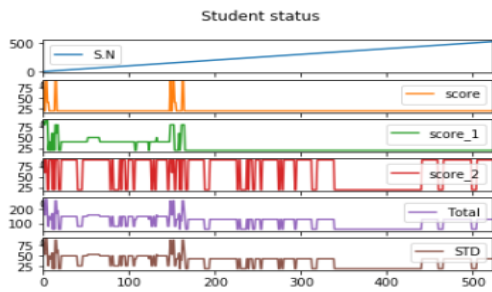
```
Out[202]:
```

	S.N	score	score_1	score_2	Total	STD	class
count	524.000000	524.000000	524.000000	524.000000	524.000000	524.000000	524.000000
mean	262.500000	21.488550	27.538168	64.770992	113.797710	37.931317	0.438931
std	151.410039	9.722629	14.137013	33.572288	42.600518	14.199519	0.816940
min	1.000000	20.000000	20.000000	20.000000	60.000000	20.000000	0.000000
25%	131.750000	20.000000	20.000000	20.000000	60.000000	20.000000	0.000000
50%	262.500000	20.000000	20.000000	90.000000	130.000000	43.330000	0.000000
75%	393.250000	20.000000	40.000000	90.000000	130.000000	43.330000	0.000000
max	524.000000	90.000000	90.000000	90.000000	270.000000	90.000000	2.000000

```
In [203]: #from pandas.tools.plotting import scatter_matrix
from pandas.plotting import autocorrelation_plot
from matplotlib import cm
```

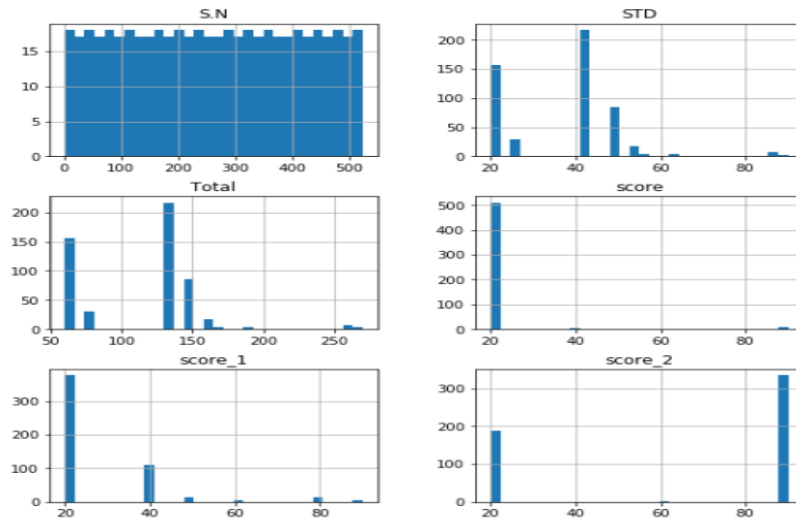
```
In [209]: ST.drop('class', axis=1).plot(subplots=True, sharex=False, sharey=False, title='Student status')
```

```
Out[209]: array([<matplotlib.axes._subplots.AxesSubplot object at 0x0000000013D3DBA8>,
<matplotlib.axes._subplots.AxesSubplot object at 0x0000000014192048>,
<matplotlib.axes._subplots.AxesSubplot object at 0x00000000148F05F8>,
<matplotlib.axes._subplots.AxesSubplot object at 0x000000001491DBA8>,
<matplotlib.axes._subplots.AxesSubplot object at 0x0000000014957198>,
<matplotlib.axes._subplots.AxesSubplot object at 0x0000000014985748>],
dtype=object)
```



```
In [210]: import pylab as pl
ST.drop('class', axis=1).hist(bins=30, figsize=(9,9))
pl.suptitle("Histogram for each numeric input variable")
plt.savefig('fruits_hist')
plt.show()
```





```
In [215]: #from pandas.tools.plotting import scatter_matrix
from pandas.plotting import autocorrelation_plot
from matplotlib import cm
#feature_names = ["score", "score_1", "score_2", "Total", "STD"]
#feature_names =
X = ST[["score", "score_1", "score_2", "Total", "STD"]]
y = ST['class']
```

```
In [220]: #from pandas.tools.plotting import scatter_matrix
from pandas.plotting import autocorrelation_plot
from matplotlib import cm
feature_names = ["score", "score_1", "score_2", "Total", "STD"]
X = ST[feature_names]
y = ST['class']
```

```
In [221]: from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, random_state=0)
```

```
In [222]: from sklearn.preprocessing import MinMaxScaler
scaler = MinMaxScaler()
X_train = scaler.fit_transform(X_train)
X_test = scaler.transform(X_test)
```

```
In [228]: from sklearn.svm import SVC
svm = SVC()
svm.fit(X_train, y_train)
print('Accuracy of SVM classifier on training set: {:.2f}'
      .format(svm.score(X_train, y_train)))
print('Accuracy of SVM classifier on test set: {:.2f}'
      .format(svm.score(X_test, y_test)))
```

```
In [229]: from sklearn.metrics import classification_report
from sklearn.metrics import confusion_matrix
pred = knn.predict(X_test)
print(confusion_matrix(y_test, pred))
print(classification_report(y_test, pred))
```

```
[[102  0  2]
 [  0  1  0]
 [  0  0 26]]
      precision    recall  f1-score   support

     0       1.00      0.98      0.99       104
     1       1.00      1.00      1.00         1
     2       0.93      1.00      0.96         26

 accuracy          0.98
 macro avg          0.98
 weighted avg          0.99
```

IV. CONCLUSION AND RECOMMENDATION

This work explores the potential of machine learning algorithms in deciding general cause of students problems and education problems, .It is found that support vector classification performance is best than that of different algorithms employed in the study. Finally, system tested using Test data Scored 98% accuracy.. This study is going to be terribly useful for the educational institutions. In future, it is doable to increase the analysis by using different clustering techniques and association rule mining for the students' dataset.

REFERENCES

1. Ambo University Legislation 2019," Ambo University, Ambo, 2019.
2. N. P. N. S. B. T. Priyanka Bhilare, "Predicting Outcome of Judicial Cases and Analysis using Machine Learning," International Research Journal of Engineering and Technology (IRJET), vol. 06, no. 03, p. 2, 2019.
3. C. M. W. L. Wenjian Wanga, "Online prediction model based on support vector machine," Elsevier, vol. 71, 2007.
4. N. Bakala, "Information Retrieval System by using Vector Space Model," INTERNATIONAL JOURNAL OF SCIENTIFIC & TECHNOLOGY RESEARCH, vol. 1, no. 1, p. 1, 2018.
5. E. G. Cyril Goutte, "A Probabilistic Interpretation of Precision, Recall and F-Score, with Implication for Evaluation," in Researc gate, Meylan, 2014.
6. K. S. J. S. Raikwal, "Performance Evaluation of SVM and K-Nearest Neighbor Algorithm over Medical Data set," International Journal of Computer Applications (0975 – 8887) , vol. 50, p. 3, 2012.
7. https://en.wikipedia.org/wiki/Ambo_University".

AUTHORS PROFILE



Naol Bakala Defersha is Lecturer and Administrative Vice Director at Ambo Univeristy, Institute of Technology. He research interest is Natural Language and Data Science.