EPRA International Journal of

Multidisciplinary Research

ENTRANTS PREDICTIVE ANALYSIS MODEL IN A PHILIPPINE STATE UNIVERSITY AND COLLEGE

Rolando Real Codilan

Faculty,
College of Computer Studies
Eastern Samar State University Main Campus,
Borongan City,
Philippines

ABSTRACT

Data mining is a process of computing models or design in a large collection of data and is used to develop predictive models. This study aims to analyze the ESSU entrance examination result from 2011-2016 of data records which subsets of Language, Reading Comprehension, Math, and Science in relation to Bachelor of Science in Computer Science (BSCS) course, build prediction analysis model on ESSU Entrance Exam, and determine the percentage or the quantity that choose BSCS course. The purpose of this study was to help in decision making and create a Predictive Analysis Model. This study used orange software for the presentation of the ESSU entrant's predictive analysis model to determine the result outcomes.

KEYWORDS *– Predictive Analysis, Entrance Exam, Orange Software*

I. INTRODUCTION

Predictive analytics has become a popular concept, it's been around for decades, more and more organizations are turning to predictive analytics to increase their bottom line. Predictive analytics uses a number of data mining, statistical algorithm and analytical techniques to bring together the management of current data to make predictions to identify the likelihood of future outcomes based on historical data.

Data Mining is the process of finding relationships and patterns within large sets of data to predict outcomes through data analysis Oliver D. Daitol et al (2012)[2]. Basically, Data Mining is mining useful information from a large set of data and using it in a more relevant manner.

Orange is an open-source software widely used the software in predicting a large set of data in data mining.

The College entrance exam is an institutionalized inclination test to measure the aggregate learning in different aptitude zones Aldous et al (2015)[1]. Because of data analysis a lot of college and university institutions, even different organizations and individual engaged in predictive analysis, especially in terms of school data records analysis to measure the future outcome of data. Students at the main stakeholders of schools/universities their performance and graduation rate are inevitable to produce a better Innovator, Entrepreneur or Global Competitors. This study aimed to predict the student's entrance examination results in a test score analysis in every category whether it could affect the first or second choice course in Eastern Samar State University (ESSU) Borongan City Philippines.

Objectives of the Study

This study aimed to design the ESSU Entrants Predictive Analysis Model that will;

1. Analyze ESSU entrance examination results from 2011-2016 of data records in test

   score in.

    1.1 Language, Reading Comprehension, Math and Science

1.2 In relation to BSCS first choice course
2. To build a predictive analysis model on ESSU's entrance exam results for the years to come.

II. METHODOLOGY
Data Collection

The data collected from the Testing and Admission Office in Eastern Samar State University was a summary report of entrance examination result for the last five years. The entrance examination result it was the only available data that this study collected.

Table 1. Summary Report on the Number of Examinees for the last five years (2011-2016).

Academic Year	Total No. of Examinees	Total No. of passers	Total No. of probationary	No. of Retakes'/applicants that did not take the Entrance Test
2011-2012	1782	705	63	4
2012-2013	562	268	47	6
2013-2014	1184	451	92	43
2014-2015	1158	852	302	36
2015-2016	2392	2045	107	19
Total	7078	4321	611	108

The table above shows the total number of examinees was not equal to the actual number of examinee gathered in individually records of students.

Predictive Analysis Model
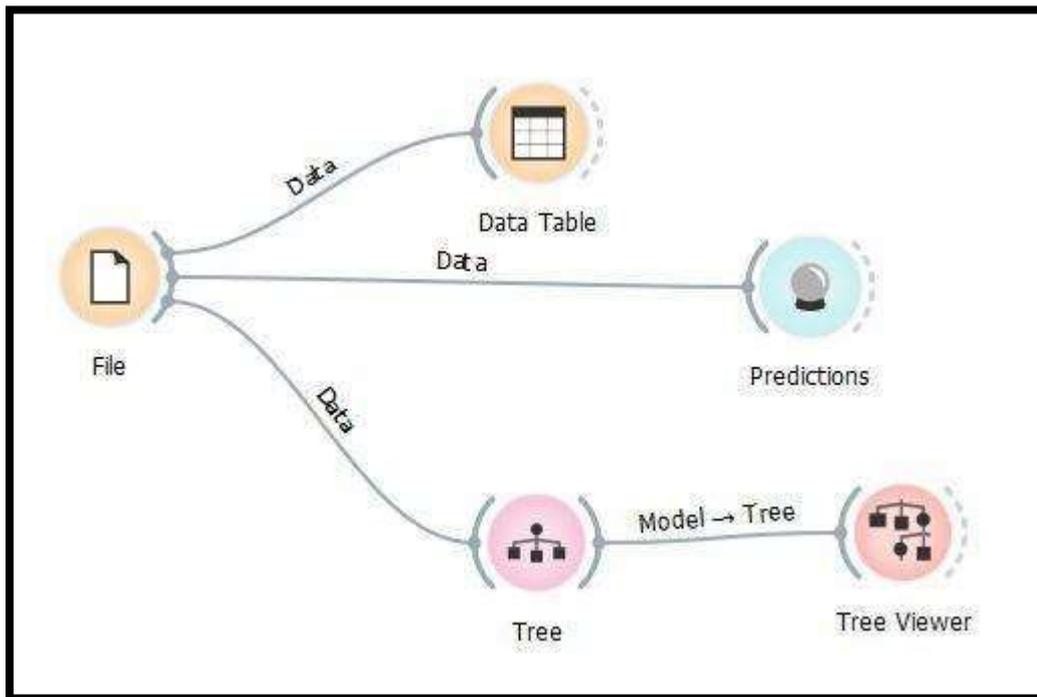
Figure 1 Predictive Analysis Model

The model above was built to understand the requirements, tested, and reworked as necessary until an acceptable prototype was achieved. This model was developed based on the currently known requirements best for predictive analysis modeling. The model aims to predict the file loaded that was the entrance examination record, store and viewed from the data table where data was stored.

Data Analysis

The data gathered from the Admission and Testing Office in Eastern Samar State University (ESSU) was encoded manually into Spreadsheet and chooses to predict and to analyze, the Year, 1st Course, Total, Remark, Reading Comprehension, Math, Science. This study used Orange as data mining tool to predict and analyze the data gathered to show the possibly predict the total number of students who will be enrolled this coming school year 2018-2019, especially in the BSCS Course,

through the percentage of each Category in language, reading comprehension, math, and science in every year; the distribution of category by year.

Research Design

This study used analytical research for which the study used the facts or data information available, and to analyze the critical evaluation of the material.

Dataset Used

Data set was a collection of data gathered from 2011 to 2016 examinees result where every single attribute of the data represents a variable and each instance has its own description. This study conducted at Eastern Samar State University (ESSU) and gathered the data from the Admission and Testing Office. The study makes the Records data into a computerized form which include 1st course, remark, year, category by; language, reading comprehension, math, science, the total number of students from the year 2011-2016 into Excel file, then imported the data into the data mining tool's which was the Orange Open Source.

Figure 2 Screenshot View of Students Entrance Record's Database

There are 7,157 rows of data from 2011-2016. The 7,157 rows represent the number of students who took the entrance examination in the past five years. For the school year 2011-2012, there were 1,792 rows of data, for 2012-2013 there were 572, for 2013-2014 there were 1,194, for 2014-2015 there were 1,187 rows of data and for the 2015-2016, there were 2,412 rows of data.

Data Mining Tool

Orange was used as the Predictive Analysis tool for experimentation and implementation. Orange was an open source data visualization machine learning and data mining toolkit. Predictive Analysis Model is done through the explorative data analysis and data visualization.

III. RESULT AND DISCUSSION
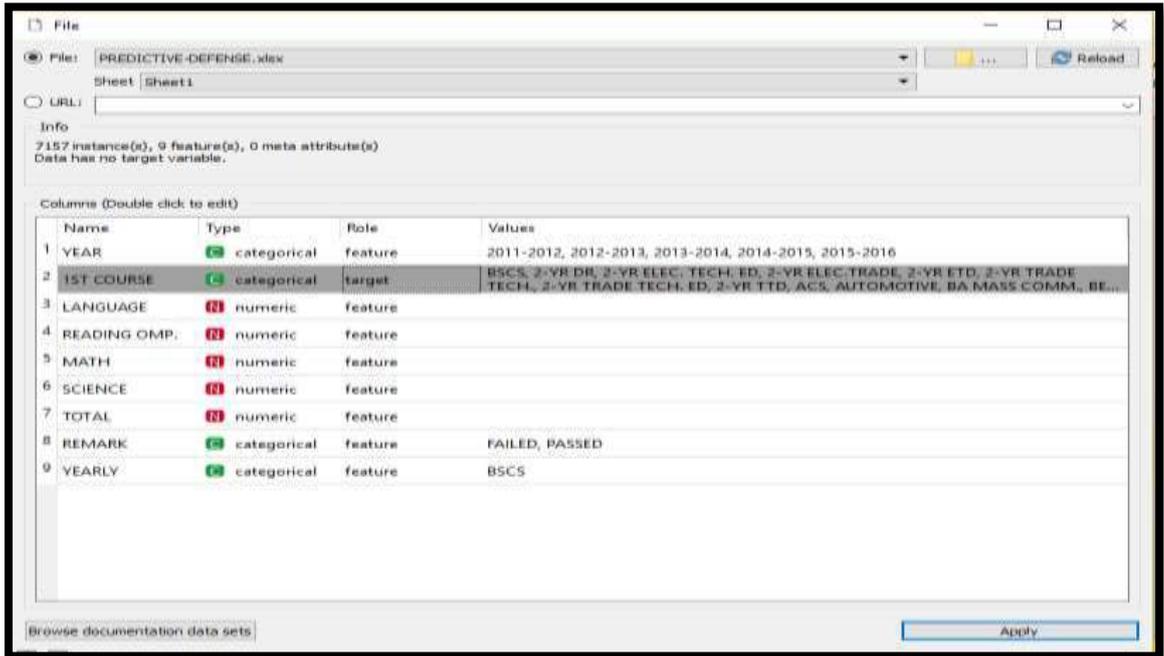
Figure 3 Screenshot View for the File Widget

Figure 3 represents the File widget used from Figure 1 that contains the following datasets discussed such as categorical for the language, reading comprehension, math, science, total, year, remark and 1st-course diagnosis. The main target of the study is the 1st course. The results contain the test scores in Language, Reading Comprehension, Math and Science in relation to the 1st-course choice.
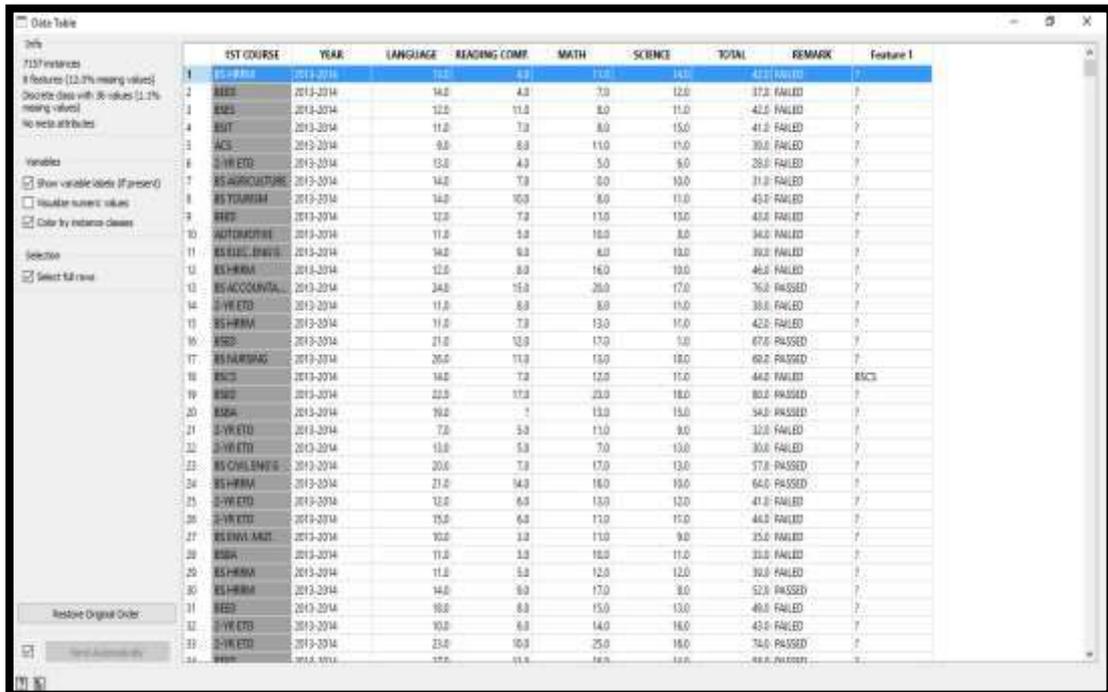
Figure 4 Screenshot for Orange Data Table

Figure 4 represents the Data table widget used from Figure 1 which represents the dataset presented by a table. The dataset was placed according to their role. The first column which represented by the 1st-course dataset filled with gray color represented the target class and the rest of the data is represented as a feature. The figure represents all Data Table widget for different months for each year.
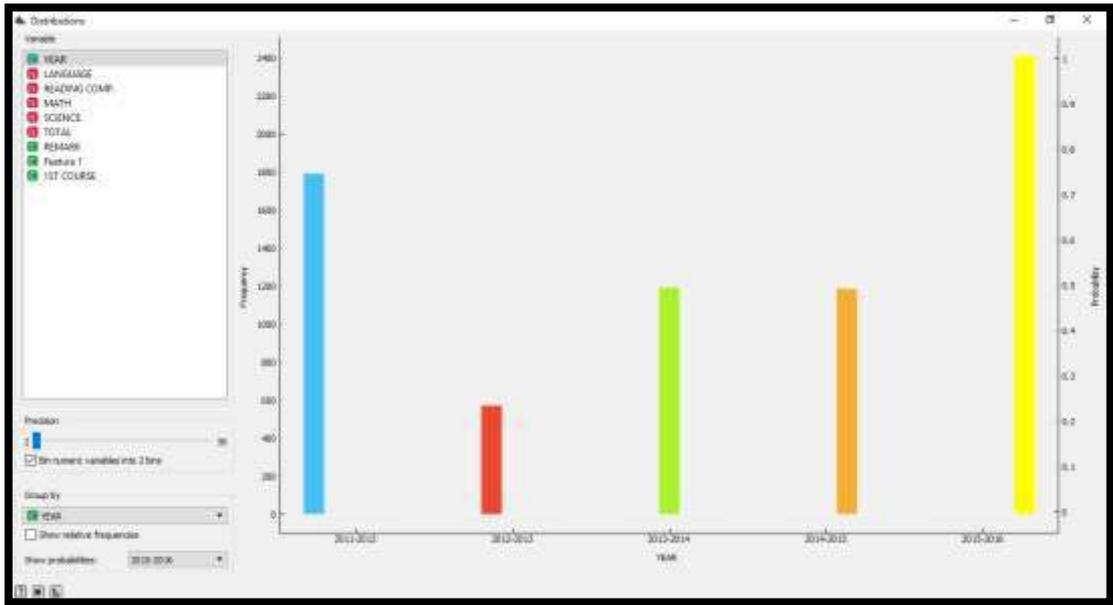
Figure 5 Screenshot for Orange Distribution

Figure 5 represents the Distribution widget used from Figure 1 which represents the distribution of five years data. The distribution represented by a graph with a different color which for each specific year.
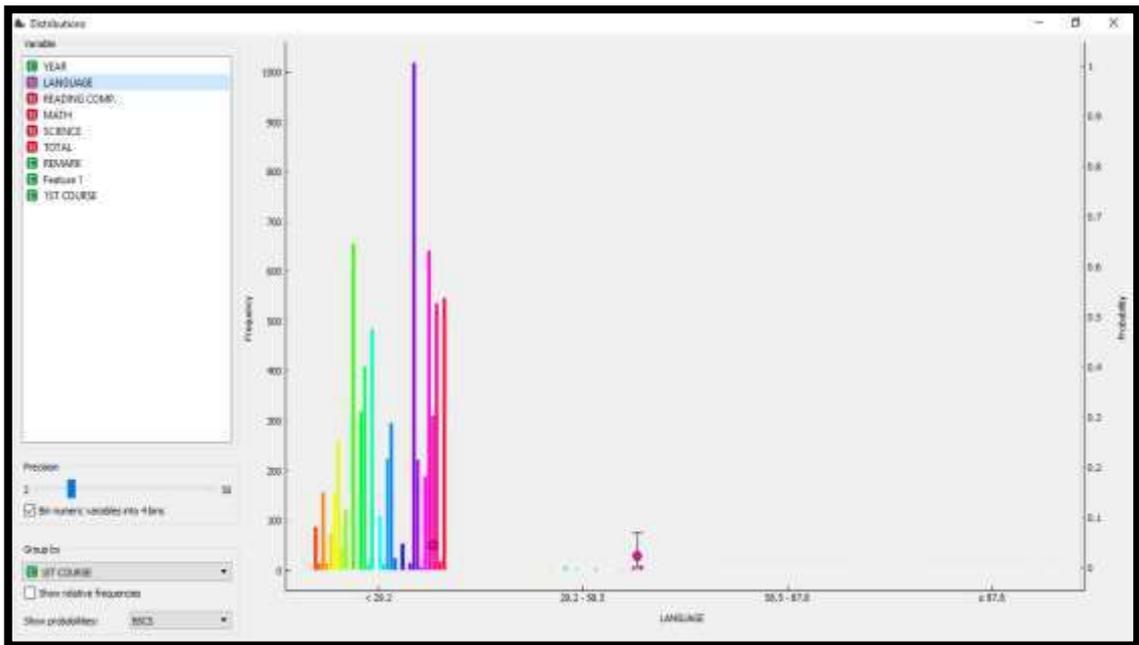
Figure 6 Distribution of 1st Course by Language

Figure 6 represents the relationship of the score in language category to the choice of courses. Based on the past five year's data in language category there are 0.44 probabilities and 0.5 percent of the student who took BS Computer Science as there course for these incoming school year.

.

Representation of Colors		
🔴 BS CIVIL ENG.	🟢 BS ELEM. EDUC.	🔵 BS BIO
🟠 BS COMP. ENG.	🟠 BS ELECTRICAL ENG.	🟢 BS ACCOUNTANCY
🟣 BSCS	🟣 BS POL. SCI.	🟢 BS AGRICULTURE
🔵 AUTOMOTIVE	🔵 BSED	🟣 BSBA
🟡 2-YR TRADE TECH.ED	🔵 BS CRIMINOLOGY	🔵 BS ENVI.SCI.
🟢 ACS	🟣 BS FISHERIES	🔵 BS HRRM
🟢 BA IN MASS COMM	🟣 BS NURSING	🟢 BSBA
🔵 BST		
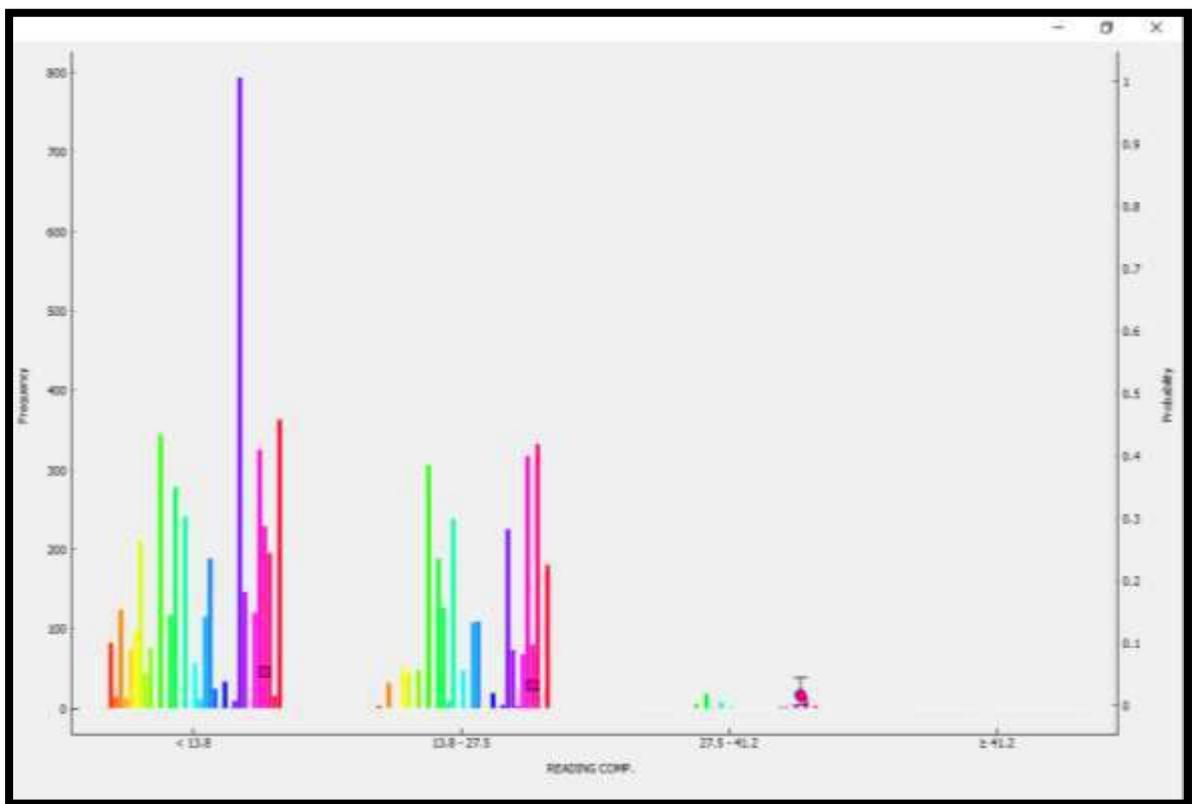
Figure 7 Distribution of 1st Course by Reading Comprehension

Figure 7 represents the relationship of the score in reading comprehension category to the choice of courses. Based on the past five year's data in reading comprehension, there are 0.68 probabilities and 0.13 percent of the student who took BS Computer Science as there course for these incoming school year.
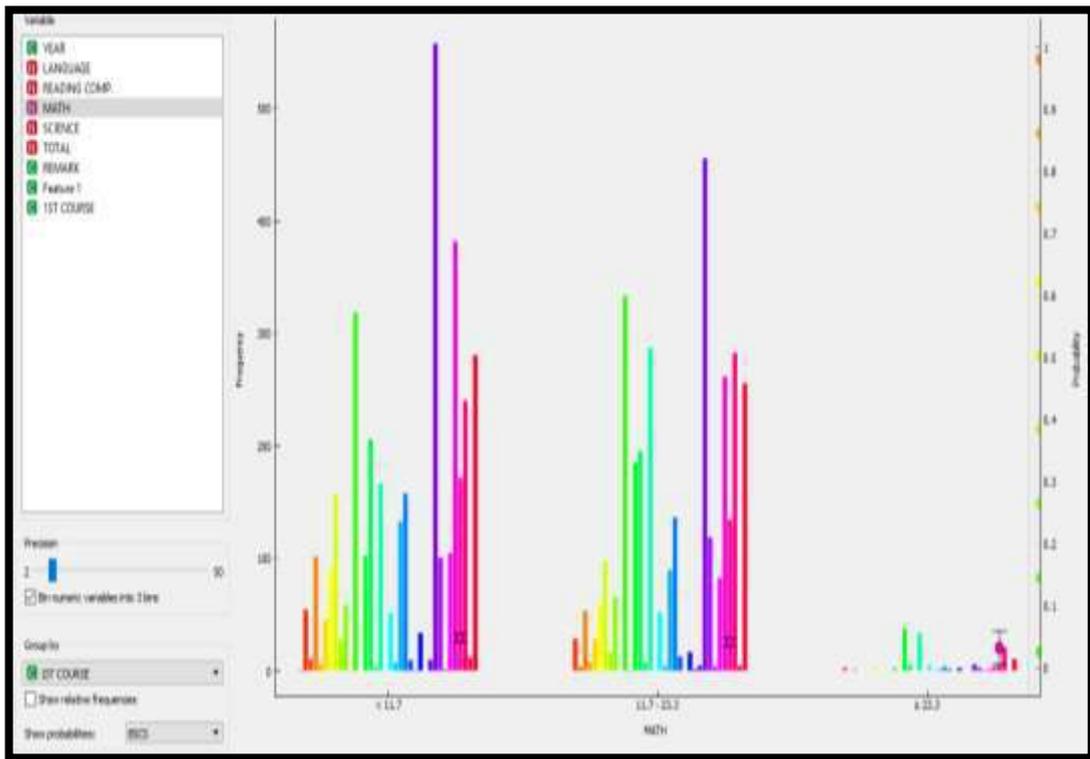
Figure 8 Distribution of 1st Course by Math

Figure 8 represents the relationship of the score in the math category to the choice of courses. Based on the past five year's data in math, there are 1.67 probabilities and 0.41 percent of the student who took BS Computer Science as there course for these incoming school year.
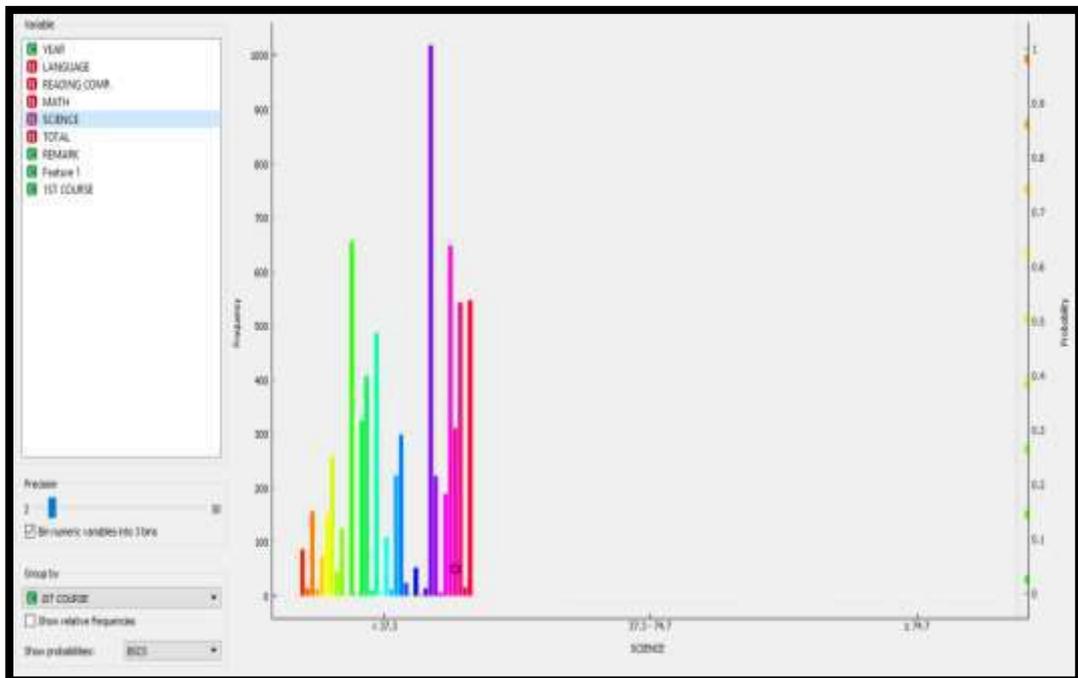
Figure 9 Distribution of 1st Course by Science

Figure 9 represents the relationship of the score in the science category to the choice of courses. Based on the past five year's data in science, there are 0.5 probabilities and 0.5 percent of the student who took BS Computer Science as there course for these incoming school year
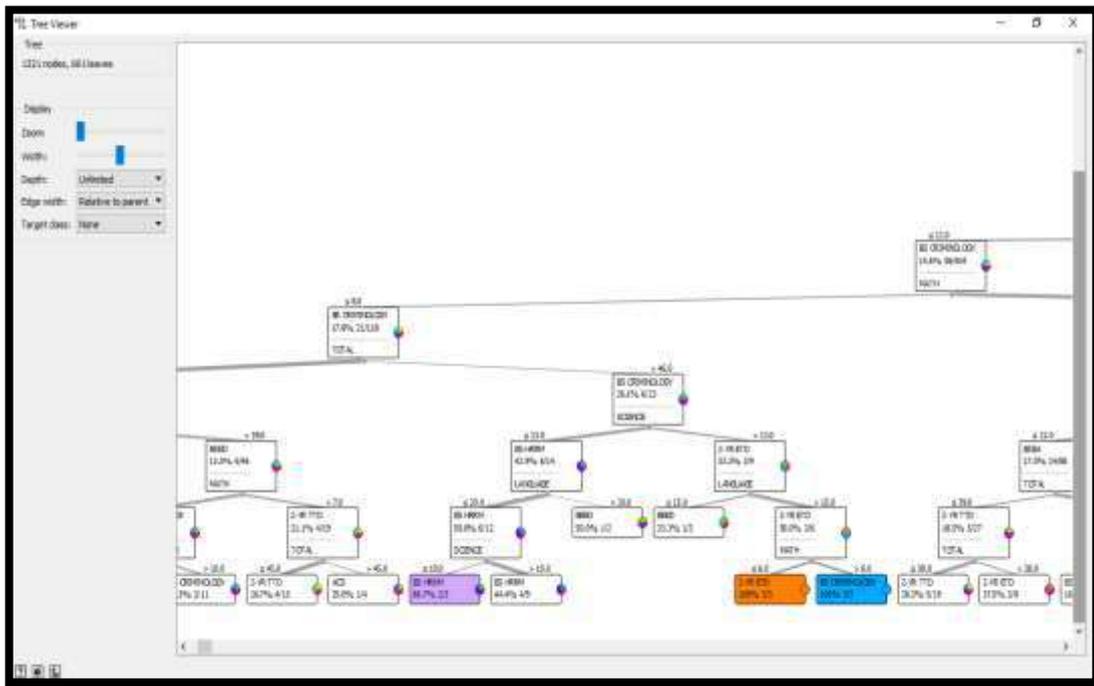
Figure 10 Screenshot View for 5 years of data

Figure 10 represents the classification tree viewer of the past five years of data. The tree viewer widget contains the result of entrance examination for the school year 2011-2016. The Tree Viewer widget used from Figure 1 that contains data by a year, ,language, reading comprehension, math, and science which is presented by a tree viewer.
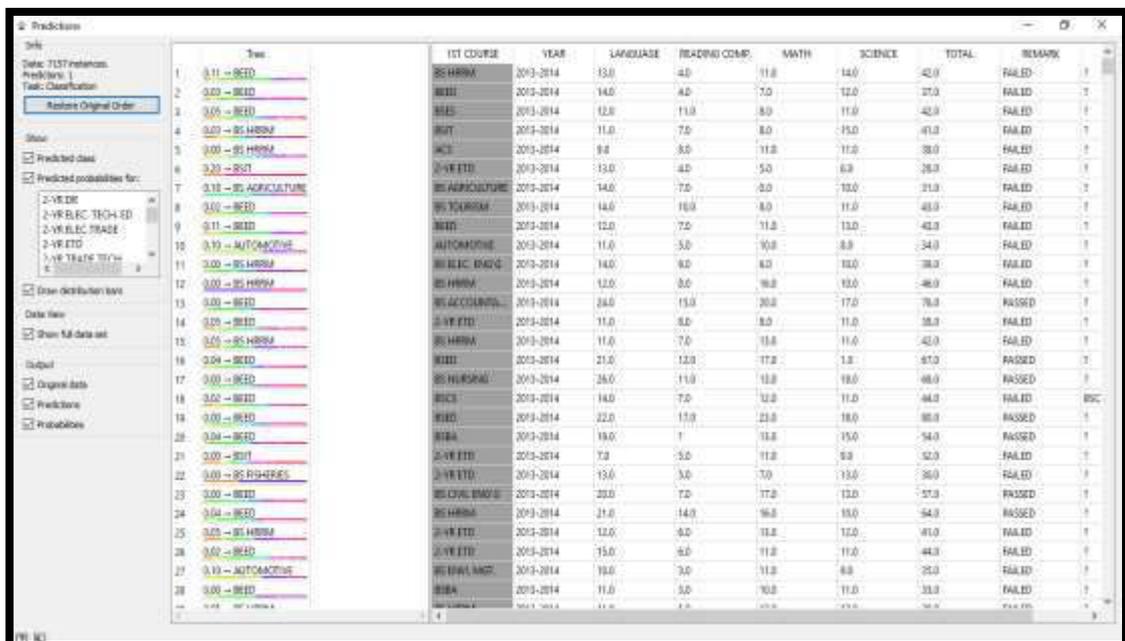
Figure 11 Prediction Analysis of 5 year's Data

Figure 11 represents the 5 years predicted probabilities for all courses in Eastern Samar State University(ESSU) for the incoming school year 2018-2019. There are 7,157 data instances in this figure.
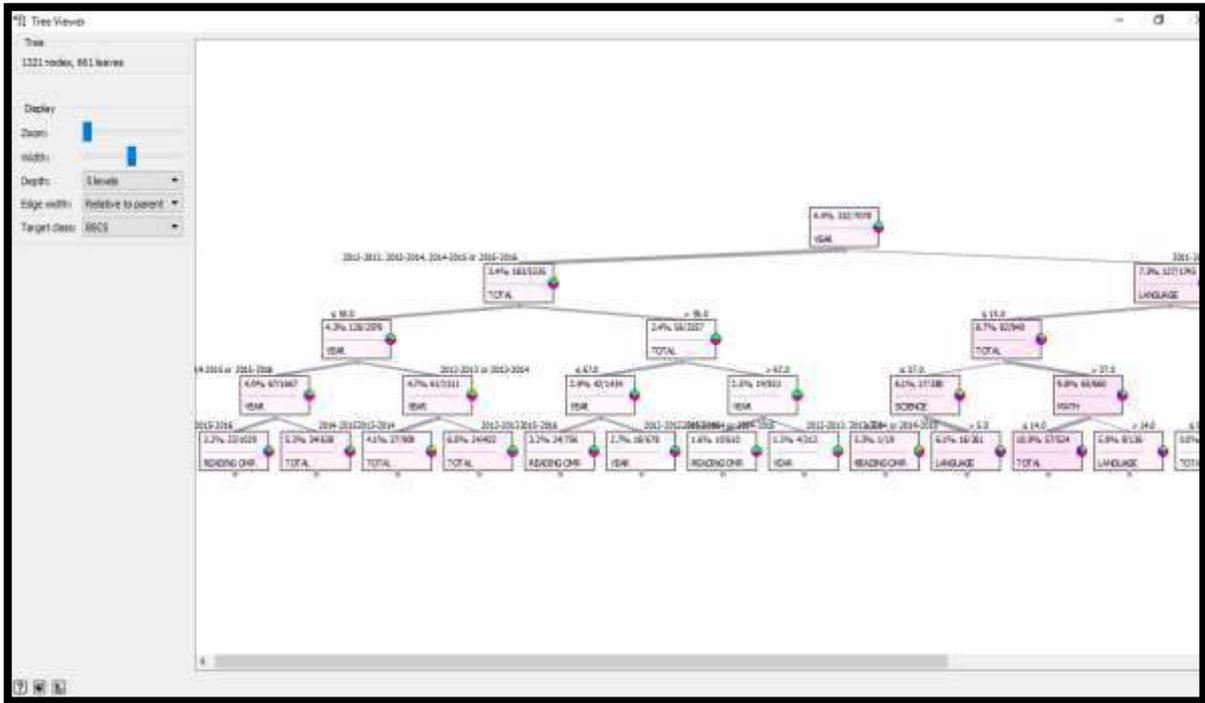
Figure 12 Screenshot for Classification Tree Viewer in BSCS Course

Figure 12 represents the classification tree viewer. It indicates that of the 310 examinees who choose BSCS as their first choice 143 who passed and 167 failed from 2011-2016. The nodes indicate the total number of examinees and the category. The leaves indicate who got the score of less than or equal or greater than the total score in every category, Language, Reading Comprehension, Math and Science.
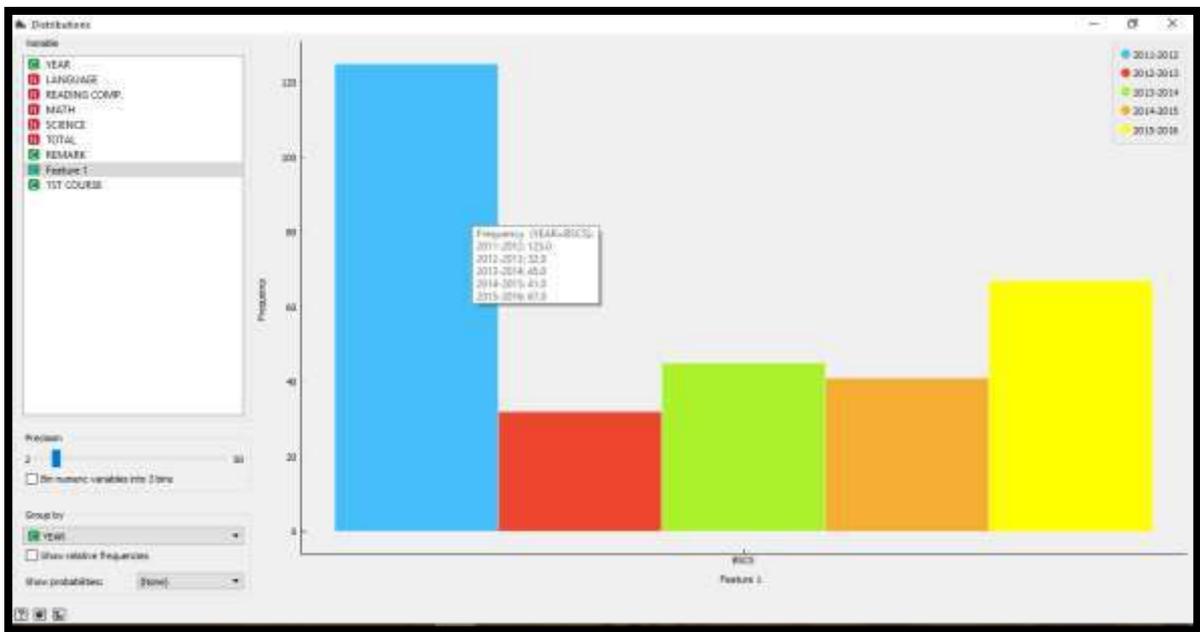
Figure 13 Graphical Distribution of 5 year's data

Figure 13 represents the graphical distribution of the past five years' data. The distribution represented by a graph with a different color for every year. The blue color is for the school year 2011-2012 in which a lot of students choose BS Computer Science Course as their first choice. The yellow represents the school year 2015-2016 which has the second highest number of student who choose BSCS as their

first choice. The green represent the school year 2013-2014 which has the third highest number of student who choose BSCS as their first choice. The orange represents the school year 2014-2015 which has the second to least number of students who choose BSCS as their first choice. Lastly, the red color represents the school year 2012-2013 which has the least number of students who choose BSCS as their first Choice.
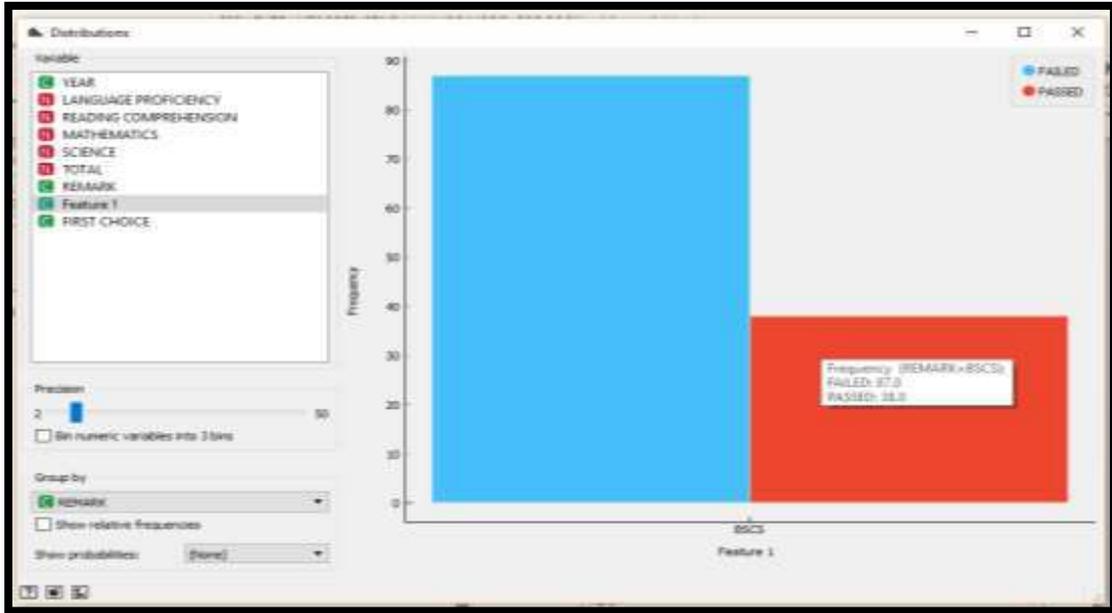
Figure 14 Graphical Distribution of School Year(2011-2012)

Figure 14 represents the distribution of passed and failed in BSCS school year 2011-2012, out of 125 students who chose BSCS as their first choice 38 passed and 87 failed

.
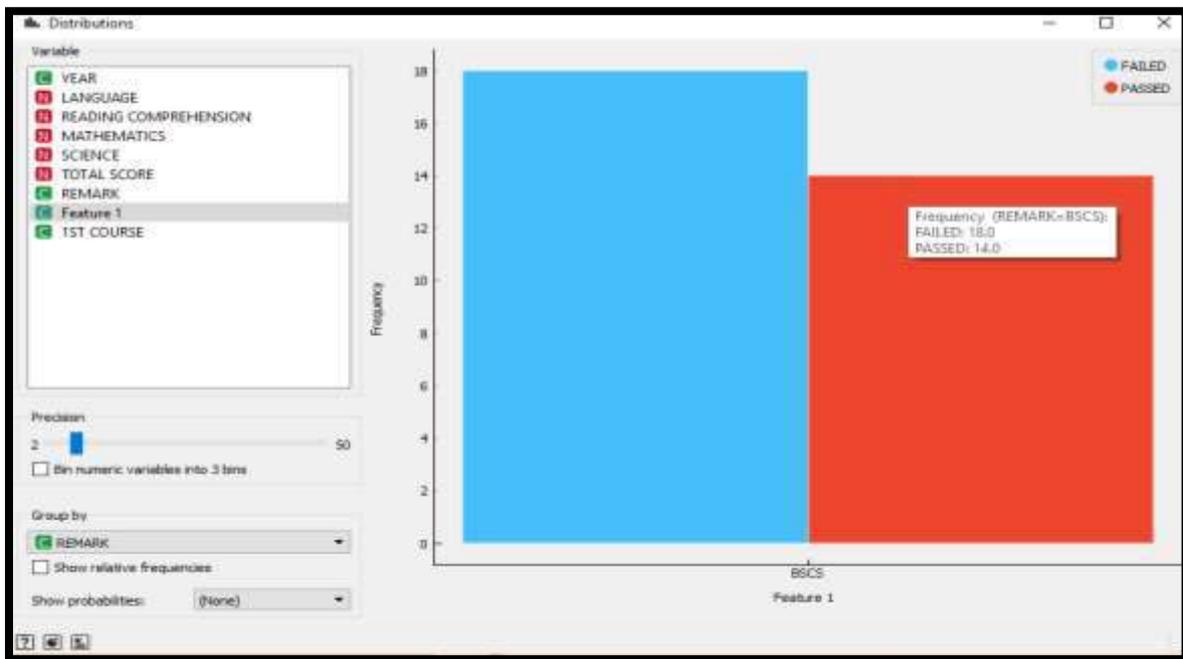
Figure 15 Graphical Distribution of School Year (2012-2013)

Figure 15 represenst the distribution of passed and failed in BSCS school year 2012-2013, out of 32 students who chose BSCS as their first choice 14 passed and18failed**.**
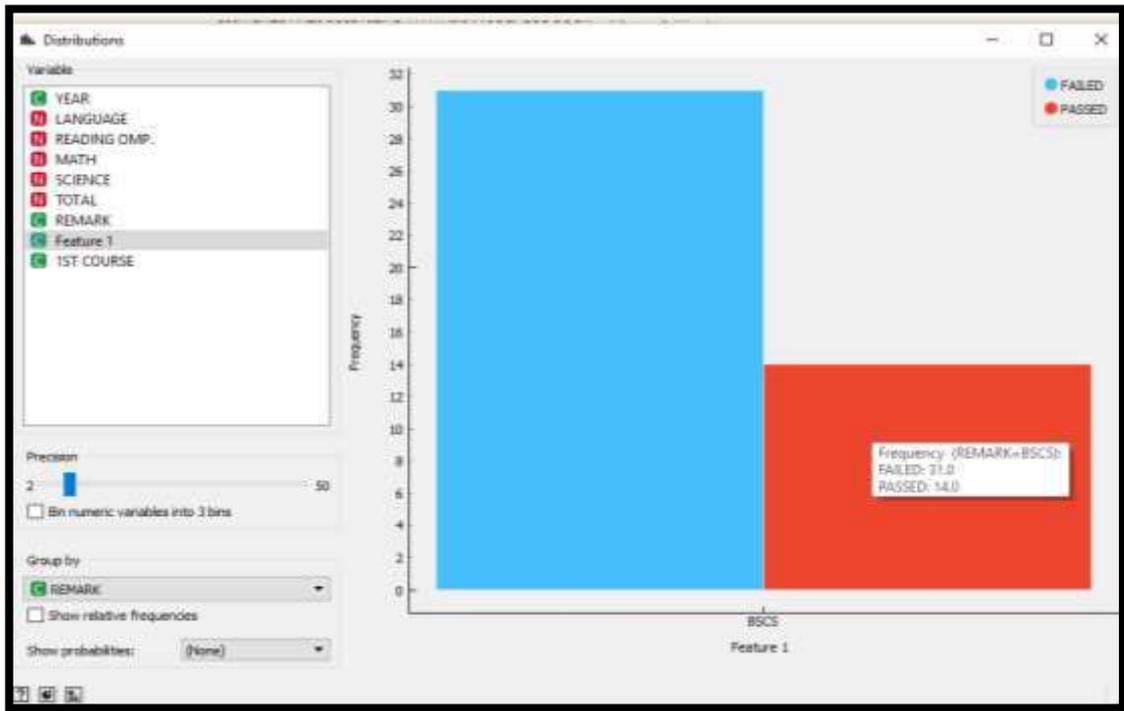
Figure 16 Graphical Distribution of School Year (2013-2014)

Figure 16 represents the distribution of passed and failed in BSCS school year 2013-2014, out of 45 students who chose BSCS as their first choice 14 passed and 31 failed.
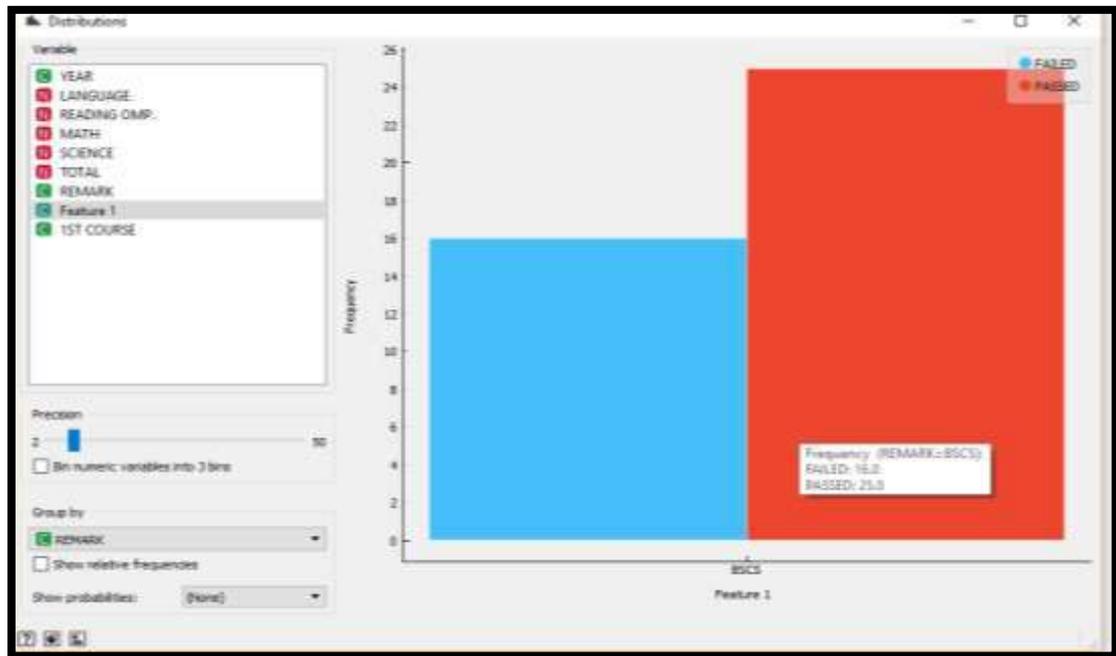
Figure 17 Graphical Distribution of School Year (2014-2015

Figure 17 represents the distribution of passed and failed in BSCS   school year 2014-2015, out of 41 students who chose BSCS as their first choice 25 passed and 16 failed.
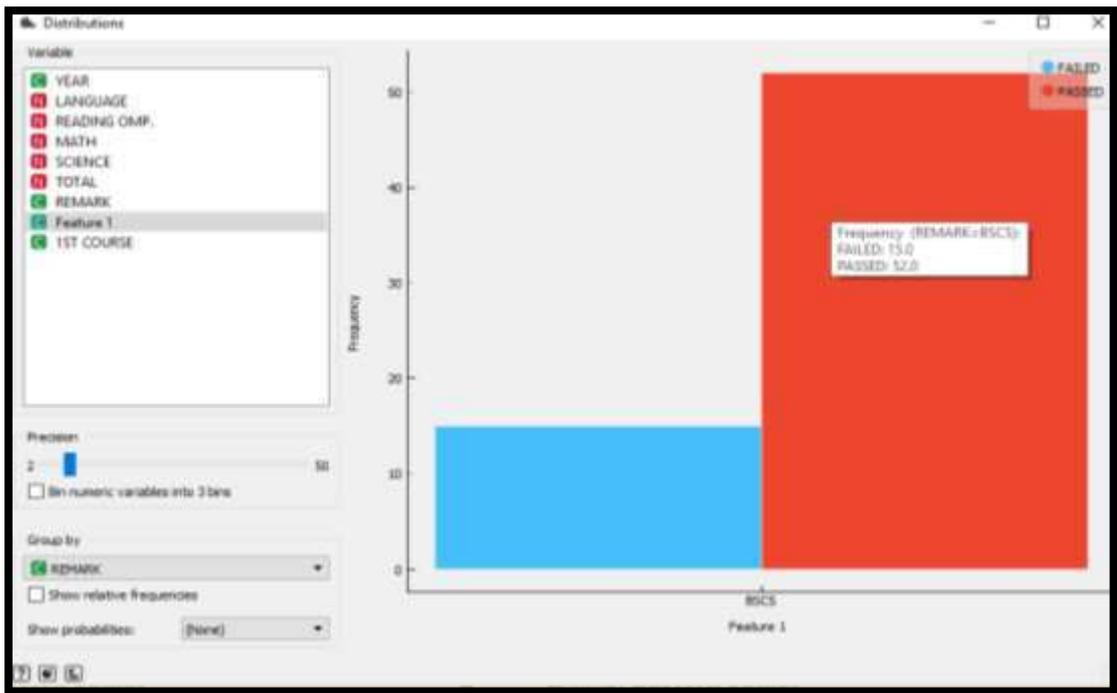
Figure 18 Graphical Distribution of School Year (2015-2016)

Figure 18 represents the distribution of passed and failed in BSCS school year 2015-2016, out of 67 students who chose BSCS as their first choice 52 passed and 15 failed.
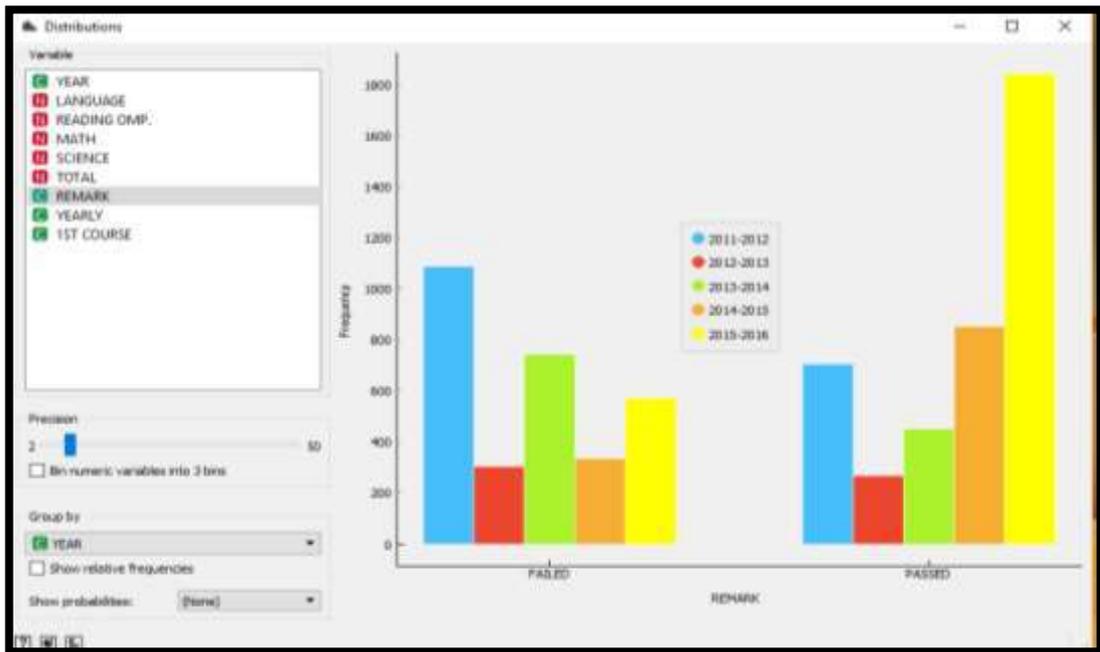
Figure 19 Graphical Distribution of 5 Year's

Figure 19 was the graphical representation of the total number of examinees in the past five years, from 2011 to 2016 of data collection.
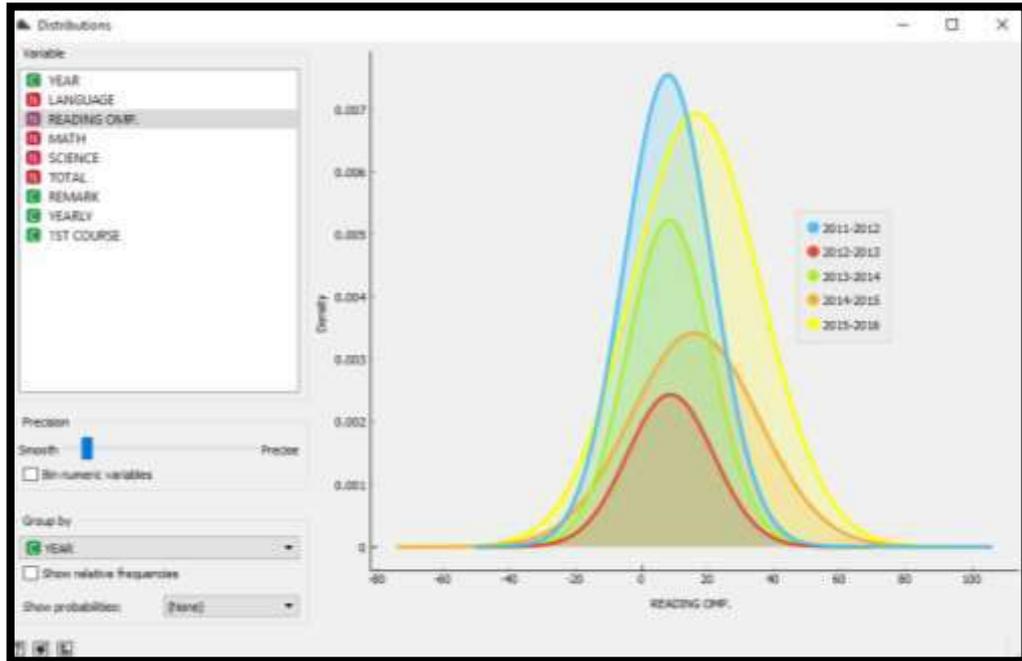
Figure 20 Graphical Analyses in Language

Figure 20 shows that the school year 2015-2016 has the highest result in the language category and 2012-2013 has the lowest result in the same category.
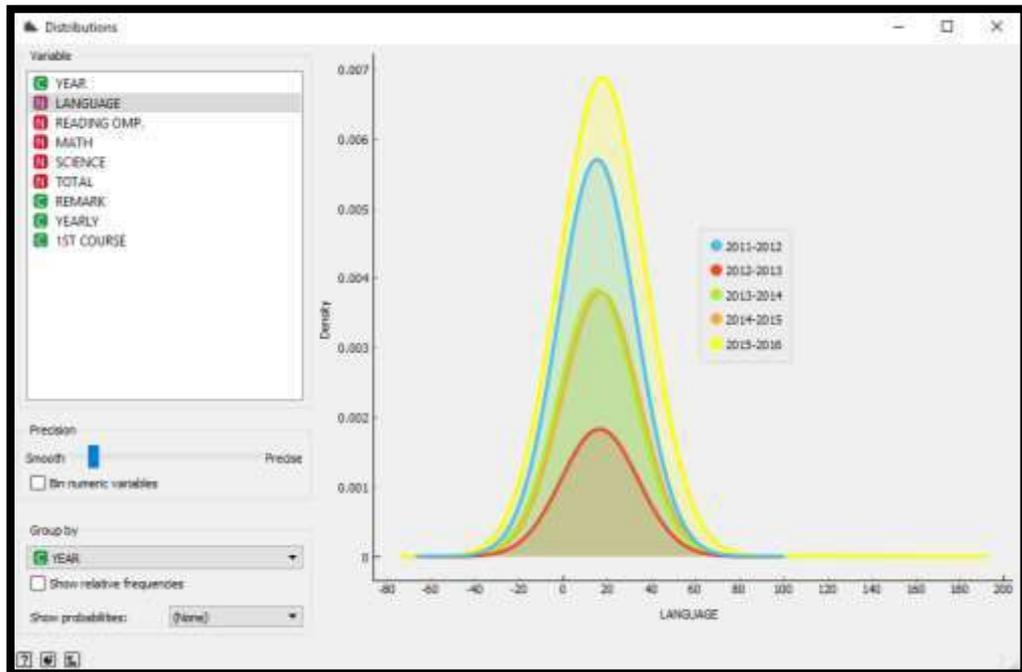
Figure 21 Graphical Distributions in Reading Comprehension

Figure 21 shows that the school year 2011-2012 has the highest result in the reading comprehension category and school year 2012 -2013 has the lowest in the same category.
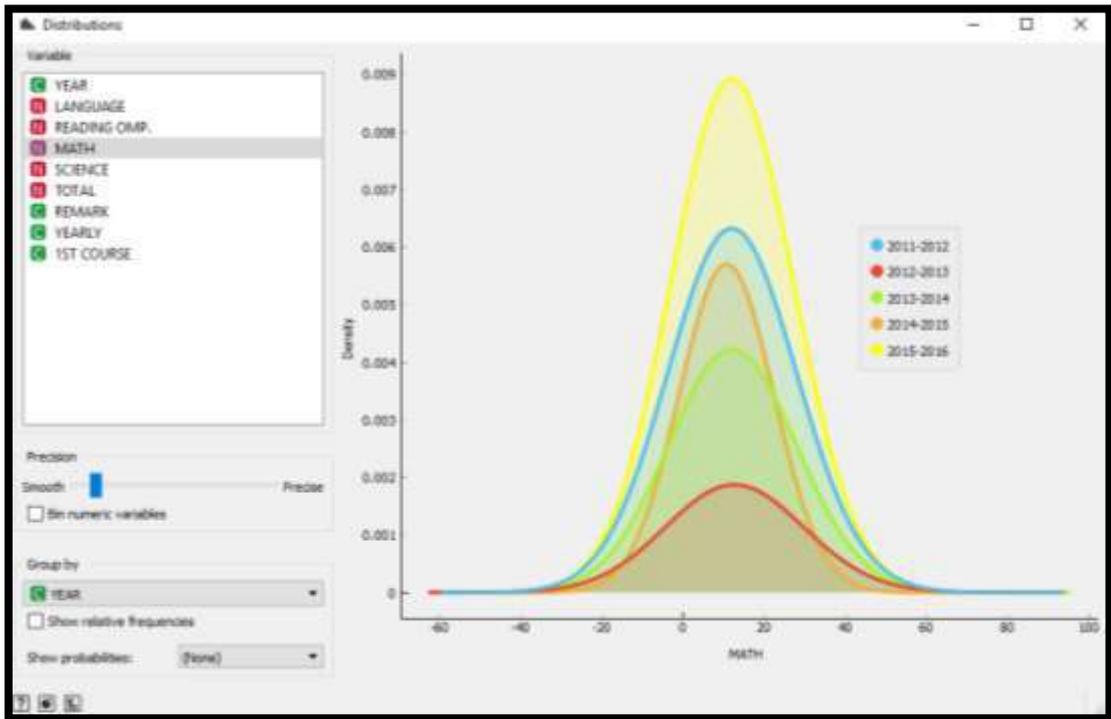
Figure 22 Graphical Distributions in Math

Figure 22 shows that the school year 2015-2016 has the highest result in the Math category and 2012-2013 got the lowest result in the same category.
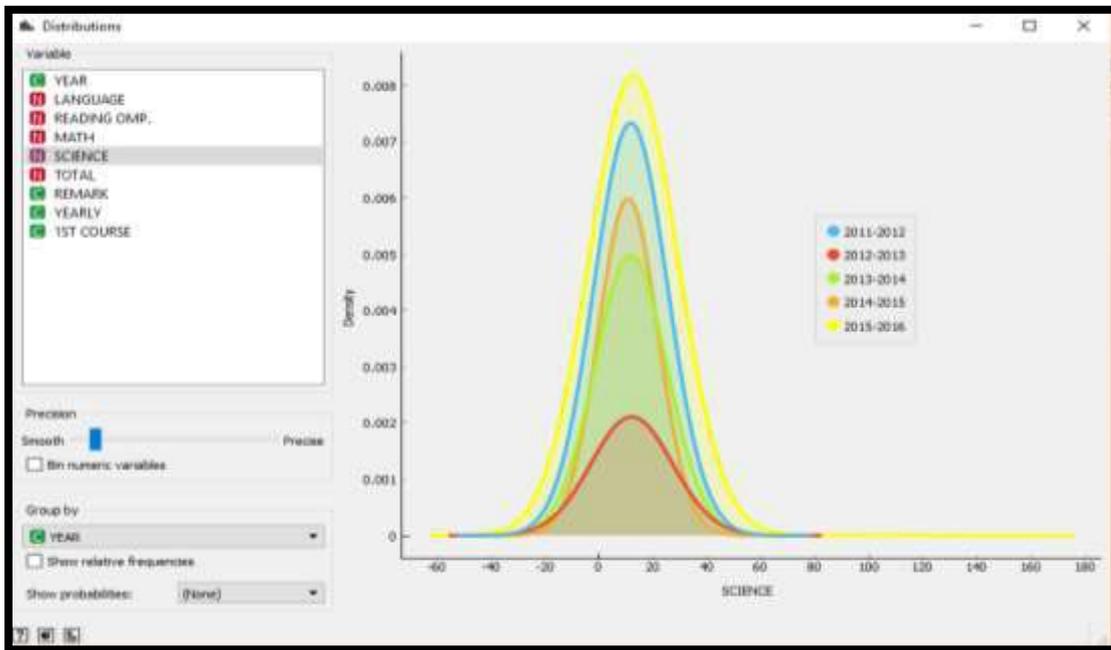
Figure 23 Graphical Analyses for Science

Figure 23 shows that the school year 2014-2015 has the highest result in the Science category and 2012-2013 got the lowest result in the the same category.

Figure 24 Prediction Analysis in BSCS

Figure 24 shows the prediction result of Class 2018-2019 BSCS Course. This study came up with this result based on the data from the school year 2011-2016. Based on the prediction there will be 108 incoming freshmen who will take BSCS Course as their first choice in School Year 2018-2019.

III. CONCLUSION

The study reveals the possible result of the school year 2018-2019 prediction analysis for each course in Eastern Samar State University. Based on the study conducted ESSU entrant predictive analysis model was able to analyze five year's data that based on test score and categories (Language, Reading Comprehension, Math, Science) targeted only BSCS as the first choice course. The Predictive analysis model was built on orange an open source software using classification tree viewer and distribution widget was used as a tool for prediction. Because of the limited data collected the prediction shows that the result depends on the test score of each student choosing BSCS as their choice course.

REFERENCES

1. Aldous Val D. Basco et al, (2015). "UPC Analysis: Predictive Analysis of the Examinees' Outcome in UPCAT for Rosales National
2. Oliver G. Daitol, Rosalie E. Cataquins, (2012). Entrance Exam Result and Performance in First Year Courses of Engineering Students: A Correlation Analysis