GEXI: COURSE PREDICTION BASED ON STUDENTS' PERSONALITY USING DATA MINING TECHNIQUE

RYAN GIL N. CABILLON
RYAN I. EDNILAN
KEVIN KHYLE M. HINOJALES
ELEONOR L. RABAÑO

A BSIT Capstone Project Submitted to the
Institute of Computing and Engineering
of Davao Oriental State University
In Partial Fulfillment of the Requirements
for the Degree

BACHELOR OF SCIENCE IN INFORMATION TECHNOLOGY

MAY 2022

Republic of the Philippines

DAVAO ORIENTAL STATE UNIVERSITY

Institute of Computing and Engineering

Guang-guang, Dahican, 8200, City of Mati, Davao Oriental

ACADEMIC INTEGRITY DECLARATION

We, **RYAN GIL N. CABILLON, RYAN I. EDNILAN, KEVIN KHYLE M. HINOJALES** and **ELEONOR L. RABAÑO** declare that this Capstone/Thesis is our own original work. Most stipulations presented herein are ours alone. Borrowed ideas are given due recognition and are properly acknowledged. With the best ability, this investigation was treated with utmost care to adhere internationally known standard/policies on academic integrity.

We attest further that this piece of academic requirements has not been submitted previously for an academic credit in this or any other courses.

RYAN GIL N. CABILLON
Researcher/Author
Date: 07/04/2022

RYAN I. EDNILAN
Researcher/Author
Date: 07/04/2022

KEVIN KHYLE M. HINOJALES
Researcher/Author
Date: 07/04/2022

ELEONOR L. RABAÑO
Researcher/Author
Date: 07/04/2022

May 2022

APPROVAL SHEET

The BSIT Capstone project herein attached entitled **"GEXI: COURSE PREDICTION BASED ON STUDENTS' PERSONALITY USING DATA MINING TECHNIQUE"**, prepared and submitted by **RYAN GIL N. CABILLON, RYAN I. EDNILAN, KEVIN KHYLE M. HINOJALES,** and **ELEONOR L. RABAÑO**, is hereby recommended for approval and acceptance.

Endorsed by:

6/24/2022

JONATHAN S. CABRERA
Adviser
Date:____________

Approved by the Panel Members.

LANIE B. LAUREANO, MIT
Chair
Date: 7/3/2022

JHON O. INOCO
Member
Date: 7/3/2022

LESLEY DIANNE BARGAMENTO
Member
Date: 7/3/2022

Accepted in partial fulfillment of the requirements for the degree in Bachelor of Science in Information Technology.

LANIE B. LAUREANO, MIT
Capstone Project Coordinator

ENGR. EMMANUEL D. BARBAS, D.Eng.
Dean, Institute of Computing and Engineering
Date: 7/4/2022

ROSELYN V. REGINO, PhD
Director for Instruction
Date: 07/06/2022

ACKNOWLEDGEMENT

First and foremost, praise and appreciation to God, the Almighty, for His showers of graces throughout the proponents' capstone project, which enabled them to successfully complete it.

The proponents would like to convey their heartfelt appreciation to the following individuals for their unwavering guidance in the fulfillment of this capstone project:

Mr. Jonatahan Cabrera, MIT, their adviser, for allowing the proponents to complete the capstone project and for providing invaluable advice during the research. His vitality, vision, genuineness, and motivation have profoundly influenced the proponents. He instructed the proponents the process for doing the research, and presented the research works as clearly as possible. It was a wonderful honor and privilege to work and learn under his supervision;

Mrs. Lanie B. Laureano, MIT, the chairman of the panel of examiners, together with the members, Mr. Jhon L. Inoco, MIT, and Ms. Lesly Dianne Bargamento, for their honest and fair comments and suggestions to improve this capstone project; and

Finally, to all the people who have directly or indirectly supported the proponents to complete the capstone project.

Thank you very much!

-The Proponents

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

Abstract

Ryan Gil N. Cabillon, Ryan I. Ednilan, Kevin Khyle M. Hinojales, and Eleonor L. Rabaño. "GEXI: Course Prediction Based on Students' Personality Using Data Mining Technique." (BSIT Capstone Project). Davao Oriental State University, May 2022.

Adviser: **Jonathan S. Cabrera**

The choice of course to enroll in a university is influenced by a student's educational experience, family background, intellectual capacity, and social spheres. This study aims to develop an application that predicts the course to enroll using the students' personalities and machine learning algorithms. To determine the personality traits by each course, the researchers used the Mini-International Personality Item Pool, a 20-item scale with four items estimating every one of the five-factor model traits. A stratified random sampling by the course was applied to identify the target respondents. The respondents are graduating students of Davao Oriental State University. This paper compares four (4) classification algorithms namely; J48, REP-Tree, Naive Bayes, and Random Tree, for predicting students' courses to enroll. The results show that the J48 classifier has the highest prediction accuracy. Thus, the J48 classifier was used to develop the web application. This study shows a broad-scale approach to help incoming first-year students decide on a course to enroll in tertiary education.

Keywords: Mini-IPIP, personality traits, machine learning algorithms, classification algorithms, student course prediction.

CHAPTER I

INTRODUCTION

1.1 Rationale of the Study

Every year, students that were bound to go to college confront the dilemma of deciding what course to take. Most teachers would acknowledge that their school aims to train students to be successful and contributing members (Honig & Hatch, 2004). Students must be equipped with the knowledge, support, and opportunity to make appropriate academic and vocational choices to attain this aim (Bower & Hardy, 2004). Choosing a college degree is undoubtedly one of the most significant decisions people can make (Edmonds, 2012). The current study's goal was to determine the most critical aspects of that choice.

Students make a series of course selections during their academic careers before each semester. Some of these selections may directly impact their future by broadening or narrowing their options for additional study and future educational and employment opportunities. Most course selection decisions shape and influence students' educational experience in intellectual and social spheres. Because early decisions influence and often define later ones, course selection decisions have a more significant impact on students' lives than they know. Knowing what criteria are utilized to choose a profession is similar to understanding what measures to choose an academic major. Deciding on an academic major is frequently the first step in establishing a professional path; the process of determining must be thoroughly explored.

Furthermore, the constraints mentioned above and concerns about the students' undetermined degree path led us to build a solution in the form of a web-based application that would assist a student in selecting a college course by answering a survey. The answers to the survey questions will assess the student's personality and generate prospective studies from Davao Oriental State University in which the student can take as major.

1.2 Objectives of the Study

This study aims to:

1. Design and develop a web-based application that will help students choose a college course based on their personality.
2. Develop a prediction model to predict the suitable course of the student based on their personalities through a personality test.
3. Compare which classifier between REP-Tree, Naive Bayes, J-48, and Random Tree has the best prediction accuracy.

1.3 Significance of the Study

The proposed project is deemed significant to the following individuals, institutions, and stakeholders:

DORsU Administration. The proposed web application can be added as an extra feature to Davao Oriental State University's present website. This tool can assist students who visit the website and seek information on enrolling.

DORsU Students. The output of this capstone project is to assist students in choosing their course based on their personalities. Incoming Davao Oriental State University students can use the proposed project in selecting a course.

Future Researchers. This study can help future researchers as this research can be used as a reference for any related investigations. Future researchers can also refer to the recommendations stated at the end of this study to improve the results gathered by the proponents and enhance the functionality of the proposed system.

1.4 Conceptual Framework

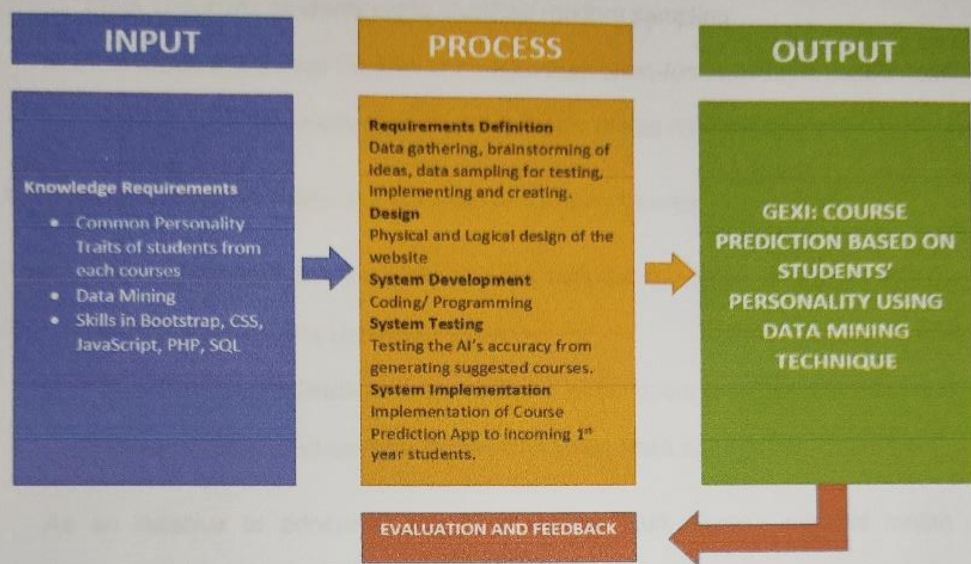

Figure 1.1 Conceptual Framework

Figure 1.1 depicts the Input Process Output (IPO) diagram of Course Prediction Based on Students' Personality Using Data Mining Techniques. The researchers and developers must complete all specifications to provide the best potential for the study. The suggested research does not need more than coding abilities, research understanding, and data mining. It must be completed inside the process immediately after the data has been obtained and processed. More observations are required to improve the system's accuracy. Feedback will be collected and used to improve the proposed application.

1.5 Scopes and Limitations

The following scope constrains the application's implementation:

1. The system is designed to predict a course based on a student's personality.

2. The personality test will be answered by 300 currently enrolled Davao Oriental State University students using stratified random sampling.
3. The researchers used the Mini-IPIP, a 20-item short-form based on the 50-item International Personality Item Pool five-factor model (Donnelan, 2006).

Although, this study's limitations are limited to a few processes:

1. The target population for gathering the data for the study is only limited to Davao Oriental State University students.
2. The GEXI Course Prediction Application's target users are incoming first-year college students and are only limited to a web-based application.

As an initiative to adhere to the Inter-Agency Task Force's minimal health regulations, this research will be confined to online questionnaires and will not require students to do face-to-face testing. Lastly, the application will not require them to provide any personal information to safeguard the students' privacy.

1.6 Definition of Terms and Study

1. Course Prediction – This application will predict a student's course in college based on their personality.
2. Incoming first-year students – Students who graduated from senior high school will start in their first year of college. They are the target users of the Course Prediction Application.
3. Mini-IPIP (International Personality Item Pool) – A five-factor model (Openness, Conscientiousness, Extraversion, Agreeableness, Neuroticism).
4. Stratified Random Sampling – The method the researchers will use to choose students from each course to take the personality test. It divides the general population into subgroups relevant to the study and ensures that the data include examples from each course.

CHAPTER II

REVIEW OF RELATED LITERATURE

2.1 Personality Traits

Trait-based theories of personality predominated the generalized personality literature throughout the 1980s and into the 1990s. Traits help describe behavioral patterns, but they are woefully inadequate for answering concerns about motivational and meaning-making components of personality. Individual differences are represented by traits, while objectives provide a method to comprehend human drives. However, neither features nor goals can express what is likely the most specific and unique part of a person and, by implication, one's personality—senses of purpose, meaning, and identity (Dunlop, 2015). Although there are various ways to think about people's personalities, Gordon Allport and other "personologists" argue that knowing their personality qualities is the best approach to grasp the distinctions between them. Personality traits are classified according to three criteria: (1) consistency, (2) stability, and (3) individual differences. (1) To have a personality characteristic, individuals must be reasonably consistent in their trait-related actions across settings. (2) Individuals with a trait are also reasonably stable in trait-related behaviors throughout time. (3) People differ in how they behave about the trait. Speaking is not a personality trait, and neither is walking on two feet—everyone does these things, and there are no individual distinctions (Edward Diener & Richard E. Lucas, 2019).

2.2 Influences on College Major Choices

According to the literature, individual characteristics, such as cognitive processing styles, learning styles, and personality factors, appear to be particularly essential in the learning process. Personality traits have been proposed as one of the key elements influencing students' performance in web-based education (Bayram et al., 2008). The value of a major is determined by both the predicted post-graduation returns and the intangible enjoyment gained during school. The popular media promotes the idea that some students avoid "useless majors." However, it is also likely that they desire to pick majors with potential employers that allow them to contribute to social impact, boosting the life outcomes of individuals with comparable economic difficulties (Liu et al., 2015). Children of educators said that their parents had a little more significant effect on their professional choices. Reviewing empirical research has also revealed that familial factors, including parental expectation and support, impact children's job goals. Expecting their children to choose a job that suits their interests, the parents highlighted the importance of education in helping their children achieve their professional goals. They pushed their children to work hard to obtain career-related abilities, generally at famous colleges, and be the best in their desired field (Liu et al., 2015). Career decisions, particularly among high school students, should be carefully considered by all students because they have far-reaching consequences in one's life. Peers affected students in many ways, including peer counseling, peer contact, peer guidance, and peer relationships. Peer counseling is a method of connecting, reacting, and assisting that explores feelings, ideas, and worries to obtain a clear understanding (Mtemeri, 2020).

2.3 Students in Shifting Courses

Most course selection decisions shape and influence students' educational experience in intellectual and social spheres. Because early decisions influence and often define later ones, course selection decisions have a more significant impact on a more meaningful life than they know at the time. Therefore, the best course selection decision-making process necessitates a substantial commitment of time and effort to gather all available information about all relevant courses and weigh all options. The problem becomes further complicated because the typical course selection process comprises a series of sequential, interconnected decisions concerning numerous courses rather than just one (Babad, 2001).

2.4 Predicting Personality

According to (Seeboth & Mõttus, 2018), the core topics of personality study are the correlations of personality traits with life outcomes, defined as occurrences that may be impacted by personality. The five major categories of the five-factor model, also known as the Big Five, illustrate personality–outcome connections (Donnellan et al., 2006): conscientiousness, extraversion, agreeableness, neuroticism, and openness.

1. Openness to Experience: curious, intelligent, and imaginative. High scorers tend to be artistic and sophisticated in taste and appreciate diverse views, ideas, and experiences.
2. Conscientiousness: responsible, organized, persevering. Conscientious individuals are highly reliable and tend to be high achievers, hard workers, and planners.
3. Extroversion: outgoing, amicable, assertive. Friendly and energetic, extroverts draw inspiration from social situations.
4. Agreeableness: cooperative, helpful, nurturing. People who score high in agreeableness are peacekeepers who are generally optimistic and trusting others.
5. Neuroticism: anxious, insecure, sensitive. Neurotics are moody, tense, and easily tipped into experiencing negative emotions.

In addition to contributing to the domain-specific facets designed to assess, they capture distinctive variation that is not shared with other domain facets and relates to separate aetiological mechanisms (Jang et al., 1998). This facet-specific variation is also connected with life outcomes; hence, adding characteristics to prediction models can improve their predictive value (Anglim & O'Connor, 2019; Christiansen & Robie, 2011; Paunonen & Ashton, 2001). For example, personality determines work happiness, professional and personal relationship success, and even preferences for distinct interfaces. Until this, users had to take a personality test to assess their personalities correctly. The capacity to forecast personality has far-reaching effects. Existing research has found correlations between personality traits and personal and professional success.

2.5 Related Systems

2.5.1 Prediction of Course Selection by Student using Combination of Data Mining Algorithms in E-Learning

The goal of a course recommender system is to anticipate the optimal combination of courses chosen by students. In this work, we show how combining a clustering algorithm—Simple K-means Algorithm—and an association rule algorithm—Apriori Association Rule—can be beneficial in a Course Recommender system. If we utilize the Apriori association rule algorithm, we must preprocess the data from the Moodle database. However, if they apply this mix of clustering and association rules, there is no need to preprocess the data. So we show this innovative technique as well as the outcome. They utilized the open-source data mining program Weka to validate the results (Aher, 2012)

2.5.2 Using Institutional Data to predict student course selections in Higher Education

Predicting which university subject a student will choose has significant quality assurance and economic implications. The ability to predict future course loads and student interests allows for more precise resource allocation, including curriculum and learning assistance and career counseling services. Prior data mining research has discovered numerous models that may be used to predict course selection based on data stored in institutional information systems. These models, however, are only intended to forecast the total number of students who may potentially enroll in a course. This earlier study did not investigate the prediction of course enrolments concerning the precise academic term and year in which students will attend such courses in the future. Furthermore, earlier models assume that all data held inside institutional information systems may be directly connected with an individual student's identification. (Ognjanovic et al., 2016).

2.6 Synthesis

The prediction of course selection predicts the optimum combination of courses by students based on the associated system mentioned above. The system demonstrates how combining a clustering technique (Simple K-means Algorithm) and an association rule algorithm (Apriori Association Rule) in a Course Recommender system may be helpful. The paper "Using Institutional Data to Forecast Student Course Selections in Higher Education" prior's data mining research has identified several models that may be used to predict course selection based on data contained in institutional information systems.

Among the related systems being presented, the "Using Institutional Data to predict student course selections in Higher Education" is close to the developed system. The associated system defines the elements that impact student course selection in order to develop a model for extracting data from student information systems and transforming it into an appropriate format. The extracted factor values are then allocated to courses provided by an institution, allowing predictions for course enrolment to be generated based on students' preferences. The "GEXI: Course Prediction Based on Students Personality" gathered data by answering the 20-item

Mini-IPIP questionnaire to 300 graduating students of Davao Oriental State University and applied data mining techniques to develop a prediction model.

CHAPTER III

MATERIALS AND METHODS

3.1 Software and Methodology

3.1.1 Software Development Life Cycle (SDLC)

The suggested framework would use a software development methodology as a broad structure. The researcher used it as a guide to follow the entire software engineering cycle. It is typically divided into six to eight steps: Planning, Requirements, Design, Build, Document, Test, Deploy, and Maintain. SDLC is a method for measuring and improving the development process. It enables a fine-grained examination of each stage of the process.

The Software Development Life Cycle succinctly describes each process involved in creating a software program. This helps to decrease waste and improve the development process's efficiency.

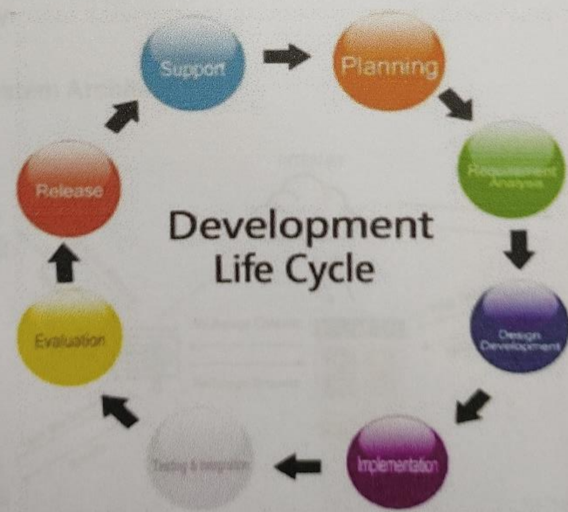

Figure 3.1 – SDLC Methodology (Rastogi, 2015)

3.1.2 Requirement Analysis

This research needs to analyze all requirements to complete the system and obligate the needs required for its development to be completed with an extensive analysis to accomplish the intended output.

3.1.2.1 Documentation of the Current System

Incoming Davao Oriental State University students must take the State College Aptitude Test (SCAST), which assesses six broad aptitude characteristics of new enrollees and transferees closely connected to their skills and abilities. This test is administered by the Guidance Counseling and Testing Center (GCTC). The university also conducts tests like the Scholastic Aptitude Test for Adults (SATA), BAR-ON Emotional Quotient Test, and the Otis-Lennon School Ability Test (OLSAT). Unfortunately, the university does not conduct a test to predict a student's course based on the student's personality.

3.1.2.2 System Architecture

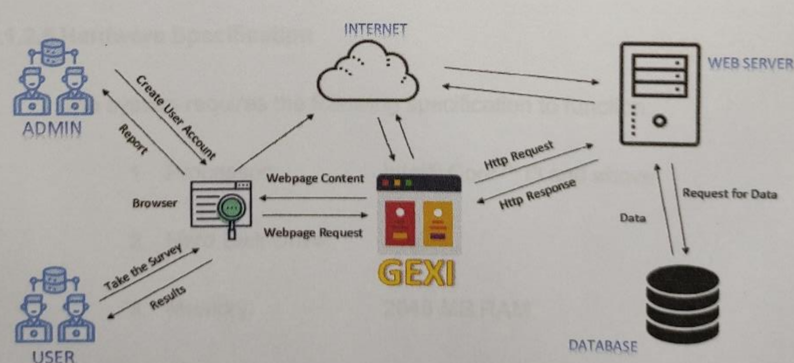

Figure 3.2 - System Architecture for GEXI: Course Prediction Application

Presented in Figure 3.2 is the whole process of the GEXI: Course Prediction Application. The main process of the system follows the administrator generating a user account for the student and receive reports from the user and non-users. Only the students with a user account can access their results from the survey.

3.1.2.3 Security

The GEXI: Course Prediction Application has a security feature for its users. Only the administrator has access to the data of the survey. The developers created a log in function with a user validation feature for the user login and administrator login. A student with a user account can only access their results from the survey.

3.1.2.4 Software Specification

The system requires the following specification to function properly.

1. Operating System: Windows 10
2. Code Editing Tool: VS Code, XAMPP
3. Database: PhpMyAdmin

3.1.2.5 Hardware Specification

The system requires the following specification to function.

1. Processor: Intel® Core™ i3 and above
2. Hard Disk Drive: 500
3. Memory: 2048 MB RAM
4. Monitor: Generic Monitor
5. Keyboard: 108 Keys

6. Mouse: Wired Optical Mouse

3.1.2.6 Development and Testing

The developers developed the system by gathering data from the survey. While waiting for the data to be gathered, the system's front-end design and certain functions for the scoring of the survey were created. These include the login function for the administrator and student of Davao Oriental University, a table that tabulates the generated answers and results from the survey, and an algorithm for the scoring and validation of the survey. After gathering the results and having the prediction algorithm, the system was finalized by implementing it to the course prediction page.

3.1.2.7 User Testing

The user may also test the website for crash and bug detection. A browser, internet connectivity, and any computer or mobile phone running Android version 6 or above are required for testing the website and mobile application. The creators conducted a questionnaire for rating and feedback purposes when testing the system. For the evaluation, the developers employed a customized questionnaire based on ISO 9126-1.

3.1.3 Design

3.1.3.1 Use Case Diagram

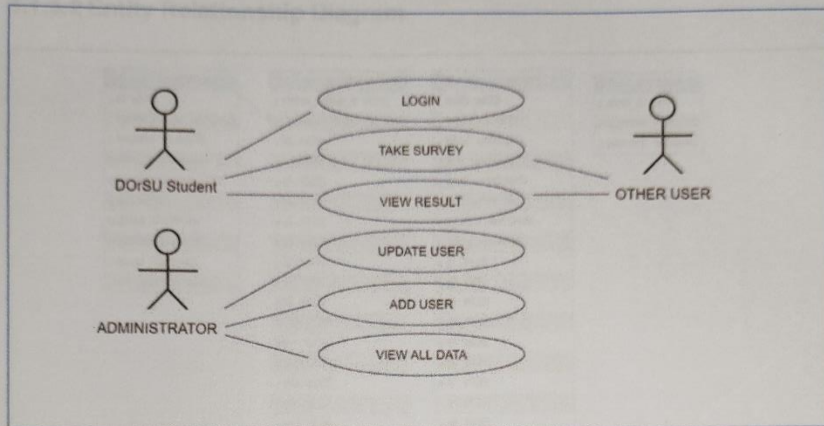

Figure 3.3 - Use Case Diagram of GEXI: Course Prediction App

Presented in Figure 3.3 is the use case diagram of the system. This indicates the limitations of the system provided for the user. For the said application, there are two types of users: (1) the administrator, and (2) the DOrSU students. Also, users who don't have an account who can only visit the application and take a survey. The administrator can view the answers for each item, and the results of the survey taken by both a student with an account and a user with no account. It is also possible for the administrator to add/create, update, and view an account of a student, specifically, students of Davao Oriental State University. A user can answer the survey without having an account. Before answering the survey, they will be asked to answer a form to save their personal information. After answering the survey, they will be redirected to the results page, showing the predicted course and the description of their scores from the big five factor domains. A student account has to log in before taking a survey and can only answer the survey once. If a student logged in without answer data, they will directly answer the survey, but if the student already has a record, the system will ask the user if he/she wishes to see its recorded results.

3.1.3.2 Entity Relationship Diagram

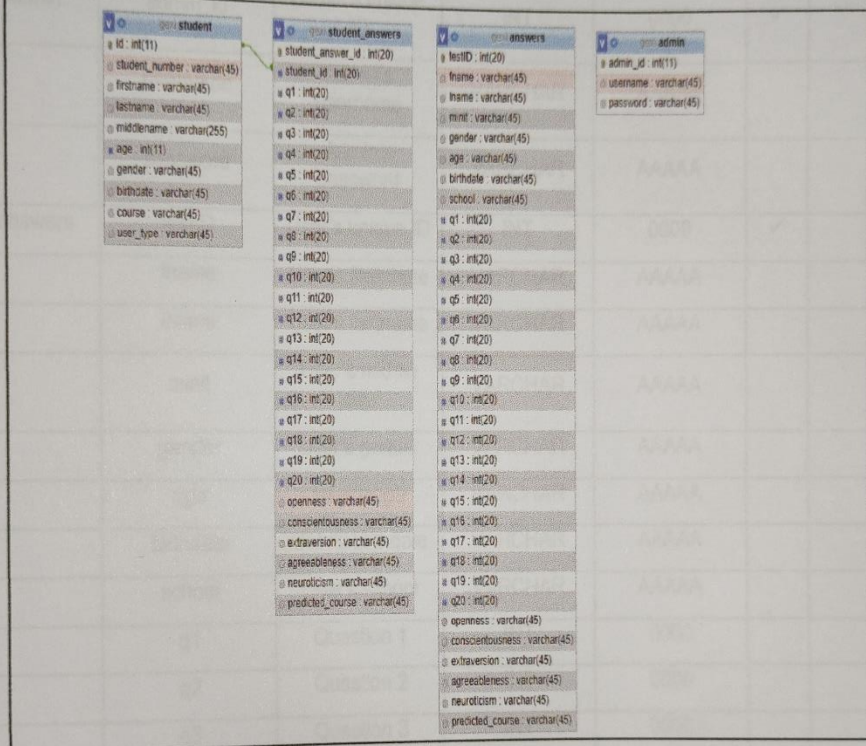

Figure 3.4 – Entity Relationship Diagram for GEXI: Course Prediction Application

Presented in Figure 3.4 is the Entity-Relationship Diagram (ERD) of GEXI: Course Prediction Application generated by PhPMyAdmin. It demonstrates the logical structure of the said application.

3.1.3.3 Data Dictionary

Presented in Table 3.1 is the Data Dictionary that provides a group of names, definitions, and attributes for data components that is utilized or captured in a database.

Table 3.1. Data Dictionary of GEXI: Course Prediction Application

TABLE NAME	ATTRIBUTE NAME	CONTENTS	TYPE	FORMAT	PK	FK
admin	admin_id	Admin's unique ID	INT	0000	✓	
	username	Admin's username	VARCHAR	AAAAA		
	password	Admin's password	VARCHAR	AAAAA		
answers	testID	User's unique ID	INT	0000	✓	
	fname	User's firstname	VARCHAR	AAAAA		
	lname	User's lastname	VARCHAR	AAAAA		
	minit	User's middle initial	VARCHAR	AAAAA		
	gender	User's gender	VARCHAR	AAAAA		
	age	User's age	VARCHAR	AAAAA		
	birthdate	User's birthdate	VARCHAR	AAAAA		
	school	User's school	VARCHAR	AAAAA		
	q1	Question 1	INT	0000		
	q2	Question 2	INT	0000		
	q3	Question 3	INT	0000		
	q4	Question 4	INT	0000		
	q5	Question 5	INT	0000		
	q6	Question 6	INT	0000		
	q7	Question 7	INT	0000		
	q8	Question 8	INT	0000		
	q9	Question 9	INT	0000		
	q10	Question 10	INT	0000		
	q11	Question 11	INT	0000		
	q12	Question 12	INT	0000		
	q13	Question 13	INT	0000		
	q14	Question 14	INT	0000		

	q15	Question 15	INT	0000		
	q16	Question 16	INT	0000		
	q17	Question 17	INT	0000		
	q18	Question 18	INT	0000		
	q19	Question 19	INT	0000		
	q20	Question 20	INT	0000		
	openness	Openness score	VARCHAR	AAAAA		
	conscientiousness	Conscientiousness score	VARCHAR	AAAAA		
	extraversion	Extraversion score	VARCHAR	AAAAA		
	agreeableness	Agreeableness score	VARCHAR	AAAAA		
	neuroticism	Neuroticism score	VARCHAR	AAAAA		
	predicted_course	Predicted Course	VARCHAR	AAAAA		
student	id	Student's unique ID	INT	0000	✓	
	student_number	Student's ID number	VARCHAR	AAAA-AAAA		
	firstname	Student's first name	VARCHAR	AAAAA		
	lastname	Student's last name		AAAAA		
	middlename	Student's middle name	VARCHAR	AAAAA		
	age	Student's age	INT	0000		
	gender	Student's gender	VARCHAR	AAAAA		
	birthdate	Student's birthdate	VARCHAR	AAAAA		
	course	Student's course	VARCHAR	AAAAA		
	user_type	Student's user type	VARCHAR	AAAAA		

student_answers	student_answer_id	Student's answer unique ID	INT	0000	✓	
	student_id	Student's ID	INT	0000		✓
	q1	Question 1	INT	0000		
	q2	Question 2	INT	0000		
	q3	Question 3	INT	0000		
	q4	Question 4	INT	0000		
	q5	Question 5	INT	0000		
	q6	Question 6	INT	0000		
	q7	Question 7	INT	0000		
	q8	Question 8	INT	0000		
	q9	Question 9	INT	0000		
	q10	Question 10	INT	0000		
	q11	Question 11	INT	0000		
	q12	Question 12	INT	0000		
	q13	Question 13	INT	0000		
	q14	Question 14	INT	0000		
	q15	Question 15	INT	0000		
	q16	Question 16	INT	0000		
	q17	Question 17	INT	0000		
	q18	Question 18	INT	0000		
	q19	Question 19	INT	0000		
	q20	Question 20	INT	0000		
	openness	Openness score	VARCHAR	AAAAA		
	conscientiousness	Conscientiousness score	VARCHAR	AAAAA		
	extraversion	Extraversion score	VARCHAR	AAAAA		
	agreeableness	Agreeableness score	VARCHAR	AAAAA		
	neuroticism	Neuroticism score	VARCHAR	AAAAA		

	predicted_cours e	Predicted Course	VARCHAR	AAAAA		

3.2 Research Design and Procedure

The process of gathering and interpreting numerical data is known as quantitative research. It may be used to discover patterns and averages, to create predictions, and to generalize results. The researchers administered a personality test to students in every course in order to assess their personalities. The participants in the study are chosen using random selection techniques. Every student of each course offered in DOrSU has an equal chance of being chosen for the sample.

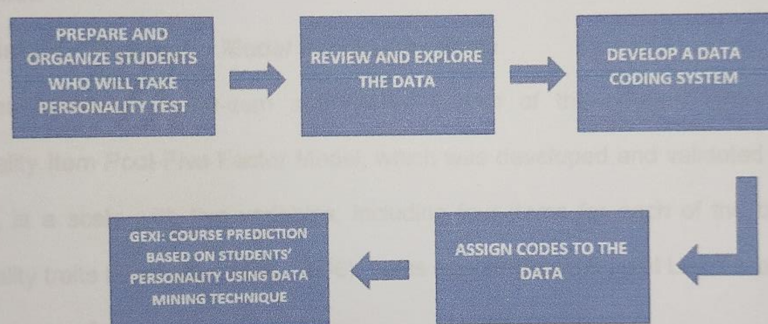

Figure 3.5 – Research Procedure

In the figure above, the flow of the research is shown. Identifying the target population who will take the personality test is the first step. The target population are students currently enrolled in Davao Oriental State University. Only 300 will be administered to answer the personality test using the Stratified Random Sampling. Stratified random sampling is a sampling method in which a population is divided into smaller sub-groups. The purpose of the stratified sampling is to get a sample population that best represents the total population being researched. The entire population is divided into homogenous groups in stratified random sampling. The formula is shown below in Figure 3.6:

$$\text{Stratified Random Sampling} = \frac{\text{Total Sample Size}}{\text{Entire Population}} \times \text{Population of Subgroups}$$

Figure 3.6 – Stratified Random Sampling Formula

The predetermined sample base is proportional to all the groups created. For example, if five groups have been created of varied sample sizes such as 10, 30, 20, 100, 60, and 80. We decide whether to choose 10% of the total population size, i.e., 300. In this case, 10% of each sample group would be chosen as the total samples to be researched. So, the numbers would be 1, 3, 2, 10, 6 and 8, and the total would be 30 samples.

3.3 Mini-IPIP Five-factor Model Personality Scale

The Mini-IPIP, a 20-item abbreviated variant of the 50-item International Personality Item Pool-Five-Factor Model, which was developed and validated in five studies, is a scale with five variables, including four items for each of the big five personality traits (Donnellan et al., 2006). Items are rated on a 5-point Likert scale with scores ranging from 1 to 5.

3.4 Sample

A total of 300 students of Davao Oriental State University from a random sample of upperclassmen undergraduates with a population size of 1,764 enrolled students took part in the research. The sample was stratified by each degree/program offered from the Institute of Computer and Engineering (ICE), the Institute of Business and Public Affairs (IBPA), the Institute of Agriculture and Life Sciences (IALS), and the Institute of Education and Teacher Training (IETT) (N = 300). Amid the pandemic, the researchers contacted the respondents by experimenters on social media (Facebook) to follow safety protocols. They were asked to fill a 3-minute questionnaire (Mini-IPIP) using a platform on the internet (Google Forms).

Table 3.2. Stratified Random Sampling Result

Course	Population Size	Sample Size
Bachelor of Elementary Education	69	12
Bachelor of Early Childhood Education	38	6
Bachelor of Special Needs Education	40	7
Bachelor of Physical Education	41	7
BPE - School Physical Ed.	2	1
BSED - Biological Science	1	1
BSED - English	39	7
BSED - Filipino	75	13
BSED - Mathematics	38	6
BSED - Physical Science	1	1
BSED - Science	28	5
BSED - TLE	1	1
BTLE - Home Economics	36	6
BTLE - Industrial Arts	62	11
Bachelor in Agricultural Technology	1	1
BS Agriculture - Animal Science	17	3
BS Agriculture - Horticulture	40	7
BS Agribusiness Management	202	33
BS Biology	48	8
BS Development Communication	29	5
BS Environmental Science	49	8
BS Nursing	50	9
BS Business Administration	221	37
BS Criminology	125	21
BS Hospitality Management	83	14
Bachelor in Industrial Technology Mgt.	162	27
BS Civil Engineering	104	17
BS Information Technology	108	17
BS Mathematics	29	5
BS Mathematics with RS	25	4
TOTAL	**1764**	300

3.5 Materials

This paper used the Mini-IPIP, a 20-item short version based on the 50-item International Personality Pool five-factor model for this study (Donnellan et al., 2006). Respondents were asked to answer each question, which was followed by a 5-point

Likert-type scale ranging from 1 (very inaccurate) to 5 (very accurate), with a neutral middle at 3 (neither accurate nor inaccurate).

The big five factors model is used to measure personality qualities that categorize students' personalities into five agents: conscientiousness, extraversion, agreeableness, neuroticism, and openness to experience. Very extraverted individuals are confident and warm rather than quiet and cautious. Agreeable individuals are well-coordinated and pleasant. Conscientious people are well-organized and exact. Neurotic people are not likely to be emotionally resilient. Finally, those who are highly open are responsive and favor novelty to routine. Using a short collection of characteristic dimensions, the Big Five can express as much variance in people's personalities as feasible (Donnellan et al., 2006).

3.6 Scoring

The Mini-IPIP is a 20-item scale that measures each of the five-factor model traits with four items. Each item is a statement that describes a certain behavior (for example, "I am the life of the party"). The participants were asked to assess how accurate this phrase is for them on a 5-point Likert-type scale. Items measuring scores from each scale were added together to obtain a total score for each of the five scales. Each OCEAN score will be calculated as the sum of four items.

Table 3.3. Mini-IPIP 20-Item Questionnaire

Big Five Domain	Text
E	Am the life of the party
A	Sympathize with others' feelings
C	Get chores done right away
N	Have frequent mood swings
O	Have a vivid imagination
E	Don't talk a lot
A	Am not interested in other people's problems
C	Often forget to put things back in their proper place
N	Am relaxed most of the time
O	Am not interested in abstract ideas
E	Talk to a lot of different people at parties
A	Feel others' emotions

C	Like order
N	Get upset easily
O	Have difficulty understanding abstract ideas
E	Keep in the background
A	Am not really interested in others
C	Make a mess of things
N	Seldom feel blue
O	Do not have a good imagination

3.7 Procedure

The researchers collected the results answered by the respondents from each question and solved the Mean, Standard Deviation, Skewness, and Kurtosis of it. The proponents have to determine whether the given data has a normal distribution - the mean, median, and mode are all the same when the data is totally normal. Furthermore, they all represent the data set's most common value. The mean, on the other hand, loses its capacity to identify the optimal central location as the data becomes skewed because the skewed data pulls it away from the normal value, which is 0. However, the median best retains its position and is not as strongly influenced by skewed values.

3.8 Data Preparation Phase

3.8.1 Data Pre-processing

This step is to prepare the data for analysis. Initially, the dataset was collected in a spreadsheet application. Feature selection has been used as a measure of dimensionality reduction to select relevant attributes (or features) from a complete set of attributes. The goal of feature selection was to select a subset of input variables by eliminating features that did not contain irrelevant or predictive information (Goswami & Chakrabarti, 2014).

3.8.2 Data Transformation

In this study, the researchers conduct online survey specifically through Google Forms. The survey data is merged in excel, verified, cleaned and analyzed using descriptive statistics by getting the mean, standard deviation, skewness, and kurtosis after collecting the data. The researchers used WEKA. It is a data mining software to create a prediction model. The data file was saved in a comma separated values (CSV) file format and later converted by WEKA to a file in Attribute Related File Format (ARFF) to simplify analysis.

3.8.3 Modeling

In the modeling phase, the modeling method was selected and applied to the dataset used in the study. This phase involves choosing the proper modeling technique, building the model, and finalizing the model (J. Kovacic, 2010).Then, model selection involves selecting techniques appropriate to the problem; refining the model whenever necessary to meet the requirements (Siraj & Abdoulha, 2009).

This approach used data mining techniques, especially classification techniques, to see how the selected variables could be predicted to be registered in GEXI. In this study, we used WEKA to simulate the output accuracy of the classifier in a more convenient way to determine which algorithm was statistically superior to the others. Classification algorithms for prediction were used, namely; J48, Random Tree, REP-Tree, and Naive Bayes. These classifiers are available in the WEKA toolkit.

Table 3.4. Selected Attributes

S/No	Attribute	Data Type	Possible Values
1	Course	Categorical	BSIT,BSBA, ...
2	Openness	Numeric	00
3	Conscientiousness	Numeric	00
4	Extraversion	Numeric	00
5	Agreeableness	Numeric	00
6	Neuroticism	Numeric	00

3.8.4 Evaluation and Deployment

During this phase, the models are evaluated to assess how well meets the business objectives and quality requirements (El, 2011). This phase consists of an iterative process of fitting different versions of the models into a set of training and testing data, each time evaluating their predictive performance. The following metrics were used to determine the performance of the model: Time taken to build the model, Kappa statistics, mean absolute Error, Root Mean Squared Error, Relative Absolute Error, Prediction number of correct predictions.

A 10-fold stratified cross-validation evaluation model was used in the final analysis. In this method, all data is divided into ten discrete sets of equal size. Stratified 10-fold cross-validation (k = 10), also known as rotation estimation, is the most common (Refaeilzadeh et al., 2020) and universal (Arlot & Celisse, 2010) evaluation models, with a more minor sample distribution variance than the hold-out cross-validation.

3.8.5 Training Experiments and Results

A two-way random partition was performed to generate the training set and the test and validation set. The training dataset was used to build the models. The number of instances the researchers used for training the data was 300. Four Classification algorithms, including; J-48, Naïve Bayes, Random Tree, and REP-Tree, were used. Table 4 shows the results obtained from the models after training.

Table 3.5. Training and Simulation Error Results

	Kappa Statistics	Mean Absolute Error (MAE)	Root Mean Squared Error (RMSE)	Relative Absolute Error (RAE)	Root Relative Squared Error (RRSE)
J48	.05	.06	.21	96.40%	118.78%
Random Tree	.04	.06	.25	96.34%	138.87%
REP-Tree	.04	.06	.18	98.27%	101.66%
Naïve Bayes	.05	.06	.18	98.23%	102.37&

CHAPTER IV

RESULTS AND DISCUSSION

4.1 Prediction Accuracy

Table 4.1. Results from each question using the Mini-IPIP Scale

Big Five Domain	Text	*M*	*SD*	Skewness	Kurtosis
E	Am the life of the party	2.8	1.18	-.02	-093
A	Sympathize with others' feelings	3.6	1.19	-.53	-0.71
C	Get chores done right away	3.8	1.10	-.82	.02
N	Have frequent mood swings	3.21	1.18	-.1	-.86
O	Have a vivid imagination	3.57	1.01	-.53	0
E	Don't talk a lot	3.11	1.19	-.20	-.93
A	Am not interested in other people's problems	2.99	1.09	-.13	-.59
C	Often forget to put things back in their proper place	2.69	1.29	.19	-1.2
N	Am relaxed most of the time	3.3	1.16	-.23	-.81
O	Am not interested in abstract ideas	2.77	1.01	-.02	-.47
E	Talk to a lot of different people at parties	2.63	1.19	.21	-.86
A	Feel others' emotions	3.52	1.13	-.62	-.29
C	Like order	3.54	1.02	-.55	-.04
N	Get upset easily	3.09	1.12	-.09	-.79
O	Have difficulty understanding abstract ideas	2.98	.82	-.10	.11
E	Keep in the background	3.21	.75	-.12	.80
A	Am not really interested in others	2.93	1.02	.01	-.44
C	Make a mess of things	2.29	1.14	.37	-.92
N	Seldom feel blue	3.07	.81	-.05	.59
O	Do not have a good imagination	2.22	1.05	.68	-.12

Table 4.2. Overall Results of the Big 5 Domain Personality Traits

	M	SD	Skewness	Kurtosis
Openness	13.6	2.34	-0.03	0.67
Conscientiousness	14.36	2.82	-0.05	-0.64
Extraversion	11.11	2.71	-0.1	-0.21
Agreeableness	13.17	3.04	-0.28	0.38
Neuroticism	11.93	2.39	0.23	0.07

The Mean, Standard Deviation, Skewness, and Kurtosis from each Big Domain Personality were shown in Table 4.2. All respondents with a sample size n = 300. Openness/Intellect has a mean of 13.6 which is neither high nor low. Conscientiousness has a mean of 14.36 which is high. Extraversion has a mean of 11.11 which is neither high nor low. Agreeableness has a mean of 13.17 which is neither high nor low. Neuroticism has a mean of 11.93 which is neither high nor low. The skewness of the 5 Big Domain Personalities are all fairly symmetrical. A fairly symmetrical distribution ranges between --0.5 and 0.5 and kurtosis is less than normal distribution or short-tailed. Statistics for the 20-item Mini-IPIP were then calculated from the relevant subset of items within this measure: M = 3.46, SD =.82; Agreeableness: M = 4.06, SD = .61; Conscientiousness: M = 3.48, SD = .76; Neuroticism: M = 2.70, SD = .61; Intellect/Openness: M = 3.64, SD = .67

After solving the overall result, the researchers gave an overview to students of Davao Oriental State University, specifically graduating students, using the five domains (John & Srivastava, 1999). In openness, students of Davao Oriental has a high rating. This indicates that the students prefer variety and being creative. They are curious about their surroundings and imaginative. Somehow some student become practical and they often find it difficult to think creatively or abstractly. In conscientiousness, the students are portrayed as disciplined, well-trained, smart, and being competent. They additionally have great motivation control, allowing them to finish tasks and achieve their goal. In extraversion, the students also have a neutral rate which is neither high nor low. This indicates that some students are sociable, enjoys being the center of attention while some dislikes being the center of attention. In agreeableness, the students also have a neutral rate which is neither high nor low. It means that some of the students can be described as soft-hearted, trusting, and well-liked while some may be perceived as suspicious, manipulative, and uncooperative. Lastly, Neuroticism. Students of Davao Oriental State University has scored low with this domain which means they doesn't worry much, calm, emotionally stable, confident, resilient, and rarely feels sad or depressed.

Table 4.3 Comparison of Evaluation Measures (Confusion Matrix)

S/ No	Classifier	Correctly Classified Instances	Incorrect Classified instances
1	J48	34	266
2	Random Tree	28	272
3	REP-Tree	41	259
4	Naïve Bayes	38	262

Generally, the correctly classified instances are mostly called the accuracy of the sample accuracy of a model. Furthermore, the REP-Tree achieved the highest performance higher than the other classifiers. These results support previous studies' argument that predictive models deserve specific evaluation methods for their performance evaluation. Moreover, from the results shown in Table 4.3, Random Tree got the lowest correctly classified instances compared to the other classifiers.

Table 4.4. Performance Measures of the Classifiers using 10-fold Cross Validation

Classifier	TP Rate	FP Rate	Recall	ROC	Speed
J48	.113	.06	.113	.538	.3 s
Random Tree	.093	.054	.093	.519	.02 s
REP-Tree	.137	.093	.137	.528	.09 s
Naïve Bayes	.127	.075	.127	.529	.01 s

In Table 4.4, the performance measure is shown by each classifier. Based on the findings of the evaluation, the REP-Tree classifier achieved a classification actual positive rate of 0.137 and a ROC of 0.538. The Response time of Naïve Bayes is 0.01 seconds. The J48 classifier achieved a classification actual positive rate of 0.113 and a ROC of 0.538. The response time of this classifier is 0.3 seconds. The Random Tree classifier achieved a classification actual positive rate of 0.093 and a ROC of 0.519 with a response of 0.02 seconds.

Table 4.5. Classifier Rankings

Classifier	Score	Ranking
REP-Tree	1.8	1st
Naïve Bayes	1.8	2nd
J-48	3.4	3rd
Random Tree	5	4th

The classifiers were ranked based on their performance, as shown in Table 4.4. In Table 4.5, REP-Tree classifier comes first with a score of 1.8 and Random Tree comes last with a score of 5. REP-Tree has the same score with Naïve Bayes. The researchers decided to make REP-Tree first in rankings since REP-Tree has the highest true positive rate. REP-Tree has the best performance but it was not used as the prediction model for this study since the sample size (300). The sample size and the variation of the performances is very small. Thus, the researchers used the J-48 classifier and used the Receiver Operating Characteristic (ROC) as a basis to use it as the prediction model for this study since it is the measure of the accuracy. A score above 0.5 would give a better prediction than random guessing. By understanding the performance given by each classifier (Shown in Table 4.4), all classifiers achieved 0.5 and above, but the J-48 classifier has the highest value than the other classifiers. Therefore, the J-48 classifier will be used as the prediction model for this study

4.2 Website Objectives

The objective of this study is to create a prediction model and design a web-based platform that will predict a student's course based on its' personality. The figures below show the developed web-based platform containing a survey questionnaire page and a results page.

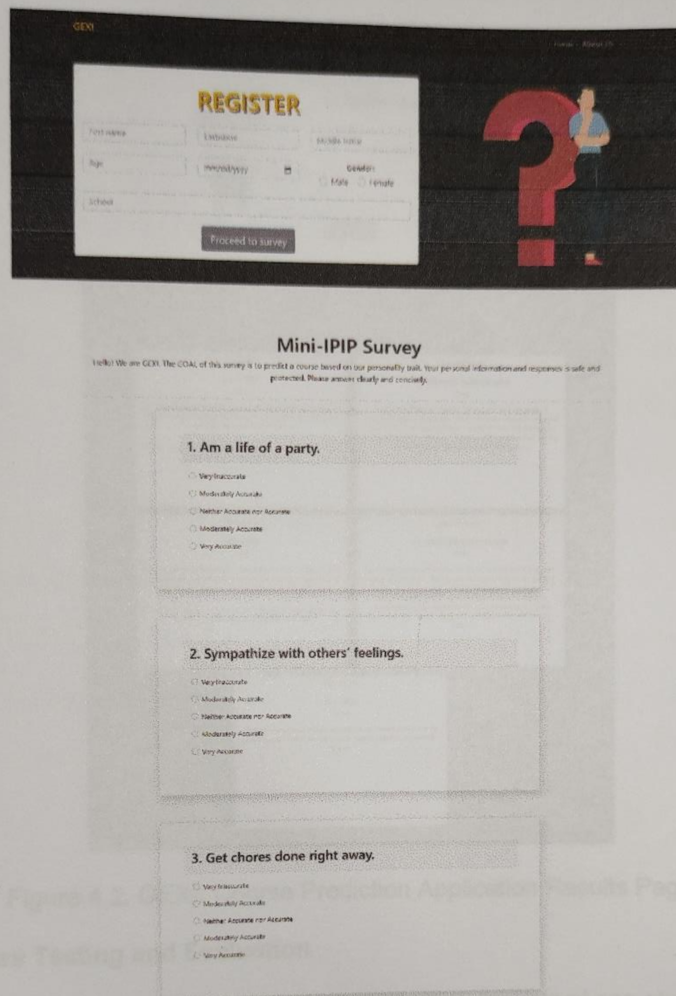

Figure 4.1. GEXI: Course Prediction Application Survey Questionnaire Page

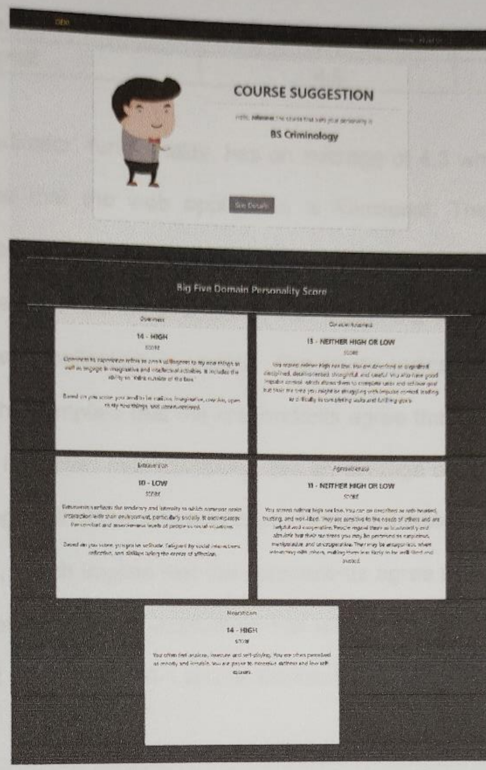

Figure 4.2. GEXI: Course Prediction Application Results Page

4.3 Software Testing and Evaluation

The proponents selected (9) respondents that consists of non-students (4) and students (5) of Davao Oriental State University for the user testing. After testing, the respondents assessed the web application through an approved survey (See Appendix B) using a Five-Point Likert Scale. The summary of the survey is shown in the table below.

Table 4.6. GEXI: CPA Evaluation Result

Indicators	Weighted Mean	Interpretation
Functionality	4.3	Agree
Reliability	4.2	Agree
Usability	4.2	Agree
Efficiency	4.8	Agree
Maintainability	4.6	Agree

Portability	4.6	Agree
Overall	**4.5**	**Agree**

The first indicator, functionality, has an average of 4.3 which implies that the respondents agree that the web application is functional. The second indicator, reliability, has an average of 4.2 which implies that the web application is reliable. The third indicator, usability, has an average of 4.2, which implies that the respondents agree that the web application is usable. The fourth indicator, efficiency, has an average of 4.8, which implies that the respondents agree that the web application is efficient. The fifth indicator, maintainability, has an average of 4.6 which implies that the respondents agree that the web application is maintainable. Lastly, portability, has an average of 4.6, which implies that the respondents agree that the web application is portable. Altogether, the web-application has an overall average of 4.5, indicating that it is compliant with ISO-9126-1 and is suitable for deployment.

CHAPTER V
SUMMARY, CONCLUSION & RECOMMENDATION

5.1 Summary

Gexi: Course Prediction Based on Students' Personality Using Data Mining Techniques is developed in the sense that this web-based application will be of great aid to those aspiring DOrSu students in searching for the most suitable course for them and for those upcoming freshmen who are reluctant on what course they would take. This application fills in as an extension between the prospective courses from Davao Oriental State University and the student, for the reason that this application would assist a student in selecting a college course by answering a survey.

The user of this web-based application was asked to answer each question, which each item is a statement describing behavior, and participants were asked to assess how accurate this phrase is for them on a 5-point Likert-type scale. After answering the survey, the researchers would collect the results answered by the respondents from each question and solve the Mean, Standard Deviation, Skewness, and Kurtosis. The researchers used WEKA as a machine learning tool to create a prediction model. The classifiers used were J-48, REP-Tree, Naive Bayes and Random tree. Among all the classifiers, the J-48 classifier achieved the best performance in prediction accuracy.

Consequently, this application is precious in the course selection of a student because just by utilizing the web-based application, it will help them to determine an adequate course for them. Lastly, this application allows the students to evaluate the best course they want to get when entering Davao Oriental State University.

5.2 Conclusion

The objective of this study was to build a prediction model to predict students' courses at DORSU based on their personalities using the Mini-IPIP personality test questionnaire. The outcomes have shown that data mining techniques can be applied to predict students' courses based on their personalities. Some people, on the other hand, contemplate the fact that anything they study will almost certainly be linked to the vocation they choose for the rest of their lives. That is why the purpose of this web-based application is to function as a guide to assist students in determining the courses they should take and to assist students in determining the courses that suit their personalities. This prediction model with the aid of the personality test will help the student find the right career and study courses. Developing and designing this application was never that easy. The developers and researchers have gone through different and challenging obstacles to develop this web-based application.

5.3 Recommendation

This application is a first step to help incoming students in finding the right course based on their personality. Therefore, there are limitations that might improve the system for future works, namely;

- Increase the sample size to improve the performance of the algorithms.
- Generate alternative courses besides from the predicted course.
- Import/export function through an excel file from the records stored in the database.
- Add a comment section for users who took the test.

In addition, many different algorithms have been developed in different data mining techniques and need to be considered in the future. This work will be improved even further by creating a prediction/recommendation system based on the results of this work.

Literature Cited

Aher, S. B. (2012). *Association Rule Mining of Classified and Clustered Data of e-Learning System*. 3(4), 10–17.

Anglim, J., & O'Connor, P. (2019). Measurement and research using the Big Five, HEXACO, and narrow traits: A primer for researchers and practitioners. *Australian Journal of Psychology*, *71*(1), 16–25. https://doi.org/10.1111/ajpy.12202

Arlot, S., & Celisse, A. (2010). A survey of cross-validation procedures for model selection. *Statistics Surveys*, *4*, 40–79. https://doi.org/10.1214/09-SS054

Babad, E. (2001). Students' course selection: Differential considerations for first and last course. *Research in Higher Education*, *42*(4), 469–492. https://doi.org/10.1023/A:1011058926613

Bayram, S., Deniz, L., & Erdoğan, Y. (2008). The role of personality traits in web based education. *Turkish Online Journal of Educational Technology*, *7*(2), 41–50.

Bower, B. L., & Hardy, K. P. (2004). From correspondence to cyberspace: Changes and challenges in distance education. *New Directions for Community Colleges*, *2004*(128), 5–12. https://doi.org/10.1002/cc.169

Christiansen, N. D., & Robie, C. (2011). Further consideration of the use of narrow trait scales. *Canadian Journal of Behavioural Science*, *43*(3), 183–194. https://doi.org/10.1037/a0023069

Donnellan, M. B., Oswald, F. L., Baird, B. M., & Lucas, R. E. (2006). The Mini-IPIP scales: Tiny-yet-effective measures of the Big Five factors of personality. *Psychological Assessment*, *18*(2), 192–203. https://doi.org/10.1037/1040-3590.18.2.192

Dunlop, W. L. (2015). Contextualized Personality, Beyond Traits. *European Journal of Personality*, *29*(3), 310–325. https://doi.org/10.1002/per.1995

Edmonds, J. (2012). *Factors influencing choice of college major: what really makes a difference?* 1–36. https://rdw.rowan.edu/cgi/viewcontent.cgi?article=1146&context=etd

Edward Diener, & Richard E. Lucas. (2019). General Psychology: General Psychology: Required Reading Required Reading. *Personality Traits*, 279–295. https://d1wqtxts1xzle7.cloudfront.net/60283871/General_Psychology_-__Required_Reading20190813-110996-103829r-with-cover-page-v2.pdf?Expires=1631502511&Signature=ZJA11PjumSnC08uj-YWExGkRdz3SYr3ZT1YEBfvOuJXRmHJjcM6SnqdfqdtuXpErEkGhfO7lFn4oZlLkzX8E5xM2Y7B8CF

Republic of the Philippines
DAVAO ORIENTAL STATE UNIVERSITY
A University of Excellence, Innovation, and Inclusion
Martinez Avenue, Dahican, 8200, Mati City, Davao Oriental

June 01, 2021

Jonathan Cabrera
Faculty – BSIT Program
Institute of Computing and Engineering

Sir,

Greetings!

We, the BSIT junior students of this institution, namely, **RYAN GIL N. CABILLON**, **RYAN I. EDNILAN**, **KEVIN KHYLE M. HINOJALES** and **ELEONOR L. RABAÑO**, would like to submit our intent to pursue with the capstone project entitled **"DORSU Web-Based Course Suggestion Application."** Thus, we wrote this letter to make a humble request that you will be our study adviser for the said project because we believe that your expertise in this field can help us complete our project.

Attached herewith is the concept paper submitted and duly approved by our capstone coordinator for your perusal.

Thank you.

Respectfully,

RYAN GIL N. CABILLON

RYAN I. EDNILAN

KEVIN KHYLE M. HINOJALES

ELEONOR L. RABAÑO

The Proponents

Noted by:

LANIE B. LAUREANO
Capstone Coordinator

RESPONSE
(Please check one for your response)

_✓_ I agree to be the adviser

____ I do not agree to be the adviser

Reason: ____________________

____

Signature

Republic of the Philippines
DAVAO ORIENTAL STATE UNIVERSITY
BACHELOR OF SCIENCE IN INFORMATION TECHNOLOGY
Guang-Guang, Dahican, City of Mati

CERTIFICATION

This is to certify that, **RYAN GIL N. CABILLON, RYAN I. EDNILAN, KEVIN KHYLE M. HINOJALES,** and **ELEONOR L. RABAÑO**, bona fide students in Bachelor of Science in Information Technology with their undergraduate thesis entitled **GEXI: Course Prediction Based on Students' Personality Using Data Mining Technique** was grammatically corrected.

This certification was issued for whatever purpose it may save them best.

Issued this 25th day of December, 2021 at Davao Oriental State University (DOrSU).

Ms. Shaye Nirza Dacles
Grammarian

grammarly Report: GEXI COURSE PREDICTION BASED ON STUDENTS' PERSONALITY USING DATA MINI...

GEXI COURSE PREDICTION BASED ON STUDENTS' PERSONALITY USING DATA MINING TECHNIQUE

by DORSU BSIT

General metrics

82,515	12,923	2550	51 min 41 sec	1 hr 39 min
characters	words	sentences	reading time	speaking time

Score

94

This text scores better than 94% of all texts checked by Grammarly

Writing Issues

368	131	237
Issues left	Critical	Advanced

Plagiarism

✓ This text seems 100% original. Grammarly found no matching text on the Internet or in ProQuest's databases.

Report was generated on Monday, May 16, 2022, 04:27 PM

Writing Issues

181 Correctness

3	Text inconsistencies
21	Incorrect noun number
1	Unknown words
24	Punctuation in compound/complex sentences
12	Faulty subject-verb agreement
9	Incorrect verb forms
12	Misspelled words
19	Wrong or missing prepositions
7	Confused words
16	Pronoun use
22	Determiner use (a/an/the/this, etc.)
6	Faulty tense sequence
1	Conditional sentences
6	Comma misuse within clauses
12	Incomplete sentences
2	Improper formatting
2	Conjunction use
1	Misuse of modifiers
1	Modal verbs
3	Misuse of semicolons, quotation marks, etc.
1	Commonly confused words

98 Clarity

27	Passive voice misuse
31	Wordy sentences

27 Unclear sentences
9 Intricate text
3 Hard-to-read text
1 Word choice

72 **Engagement**
71 Word choice
1 Monotonous sentences

17 **Delivery**
9 Incomplete sentences
7 Potentially sensitive language
1 Inappropriate colloquialisms

Unique Words

Measures vocabulary diversity by calculating the percentage of words used only once in your document

14%
unique words

Rare Words

Measures depth of vocabulary by identifying words that are not among the 5,000 most common English words.

40%
rare words

Word Length

Measures average word length

3.9
characters per word

Sentence Length

Measures average sentence length

5.1
words per sentence

Document Information

Analyzed document	GEXI Course Prediction Based on Students Personality using Data Mining Technique.docx (D140988980)
Submitted	6/21/2022 4:09:00 AM
Submitted by	Cristian D. Jael
Submitter email	jaelcristian@gmail.com
Similarity	0%
Analysis address	jaelcristian.doscst@analysis.ouriginal.com

Sources included in the report

Entire Document

GEXI: COURSE PREDICTION BASED ON STUDENTS' PERSONALITY USING DATA MINING TECHNIQUE
RYAN GIL N. CABILLON RYAN I. EDNILAN KEVIN KHYLE M. HINOJALES ELEONOR L. RABAÑO
A BSIT Capstone Project Submitted to the Institute of Computing and Engineering of Davao Oriental State University In Partial Fulfillment of the Requirements for the Degree
BACHELOR OF SCIENCE IN INFORMATION TECHNOLOGY
MAY 2022
Republic of the Philippines DAVAO ORIENTAL STATE UNIVERSITY Institute of Computing and Engineering Guang-guang, Dahican, 8200, City of Mati, Davao Oriental
ACADEMIC INTEGRITY DECLARATION
We, RYAN GIL N. CABILLON, RYAN I. EDNILAN, KEVIN KHYLE M. HINOJALES and ELEONOR L. RABAÑO declare that this Capstone/Thesis is our own original work. Most stipulations presented herein are ours alone. Borrowed ideas are given due recognition and are properly acknowledged. With the best ability, this investigation was treated with utmost care to adhere internationally known standard/policies on academic integrity.
We attest further that this piece of academic requirements has not been submitted previously for an academic credit in this or any other courses.
RYAN GIL N. CABILLON Researcher/Author Date:_______________
RYAN I. EDNILAN Researcher/Author Date:_______________
KEVIN KHYLE M. HINOJALES Researcher/Author Date:_______________
ELEONOR L. RABAÑO Researcher/Author Date:_______________
May 2022 APPROVAL SHEET
The BSIT Capstone project herein attached entitled "GEXI: COURSE PREDICTION BASED ON STUDENTS' PERSONALITY USING DATA MINING TECHNIQUE", prepared and submitted by RYAN GIL N. CABILLON, RYAN I. EDNILAN, KEVIN KHYLE M. HINOJALES, and ELEONOR L. RABAÑO, is hereby recommended for approval and acceptance.
Endorsed by:
JONATHAN S. CABRERA Adviser Date:_____________
Approved by the Panel Members:
LANIE B. LAUREANO, MIT Chair Date:_____________
JHON O. INOCO Member Date:_____________
0
LESLEY DIANNE BARGAMENTO Member Date:_____________
0
Accepted in partial fulfillment of the requirements for the degree in Bachelor of Science in Information Technology
LANIE B. LAUREANO, MIT Capstone Project Coordinator
ENGR. EMMANUEL O. BARBAS, D.Eng. Dean, Institute of Computing and Engineering Date:_______________
ROSELYN V. REGINO, PhD Director for Instruction Date:_______________
ACKNOWLEDGEMENT First and foremost, praise and appreciation to God, the Almighty, for His showers of graces throughout the proponents' capstone project which enabled them to successfully complete it. The proponents would like to convey their heartfelt appreciation to the following individuals for their unwavering guidance in the fulfillment of this capstone project: Mr. Jonathan Cabrera, MIT, their adviser, for allowing the proponents to complete the capstone project and for providing invaluable advice during the research. His vitality, vision, genuineness, and motivation have profoundly influenced the proponents. He instructed the proponents the process for doing the research, and presented the research works as clearly as possible. It was a wonderful honor and privilege to work and learn under his supervision; Mrs. Lanie B. Laureano, MIT, the chairman of the panel of examiners, together with the members, Mr. Jhon L. Inoco, MIT, and Ms. Lesly Dianne Bargamento, for their honest and fair comments and suggestions to improve this capstone project; and Finally, to all the people who have directly or indirectly supported the proponents to complete the capstone project. Thank you very much! -The Proponents

RYAN GIL N. CABILLON
Limatoc St., Brgy. Matiao
City of Mati, Davao Oriental
+639984925369
cabillonryangil@gmail.com

PERSONAL DATA

Date of Birth:	March 31, 1999
Age:	23
Place of Birth:	Quezon City, Manila
Sex:	Male
Height:	5'7
Weight:	75 kg
Civil Status:	Single
Nationality:	Filipino

SKILLS AND INTEREST

- ✓ Computer Skills
- ✓ Programming Skills
- ✓ Database Management
- ✓ Video Editing
- ✓ Decision Making
- ✓ Problem Solving

EDUCATIONAL BACKGROUND

TERTIARY	Bachelor of Science in Information and Technology Davao Oriental State University Guang-guang, Dahican, City of Mati Davao Oriental 2018-2022
SECONDARY	The Seed Montessori School No. 4 Samonte Street, Capitol District, Quezon City Manila 2011-2015
PRIMARY	The Seed Montessori School No. 4 Samonte Street, Capitol District, Quezon City Manila 2005-2011

WORK EXPERIENCE

YOGIYO CHICKEN Korean Restaurant
4A EK Building, 100 Holy Spirit Dr, Lungsod Quezon, 1127 Kalakhang Maynila
Food server, Kitchen staff, Delivery
(August 2017 – November 2017)

KANDLE CAFE
1103, 113 Mother Ignacia Ave, Diliman, Quezon City, 1103 Metro Manila
All around crew
(November 2017 – March 2018)

VOICE ACTOR (FREELANCE)

ABS-CBN (Media Company)
ABS-CBN Broadcast Center, Sgt. Esguerra Avenue cor Mother Ignacia Street, Diliman, Quezon City 1103 Philippines

VIVA Studio Sct. Madriñan
Scout Madriñan, Quezon City

Beginnings at Twenty-Plus Inc.
6th Floor, Don Chua Lamko Bldg., 100 LP Leviste corner HV Dela Costa Sts., Salcedo Village 1227

(2015-2018)

RYAN I. EDNILAN
Camansi, Brgy. Badas
City of Mati, Davao Oriental
+639978015127
ednilan1992@gmail.com

PERSONAL DATA

Date of Birth:	July 10, 1992
Age:	29
Place of Birth:	Santo Rosario, San Isidro Davao Oriental
Sex:	Male
Height:	5'5
Weight:	70 kg
Civil Status:	Single
Nationality:	Filipino

SKILLS AND INTEREST

- ✓ Computer Repair
- ✓ Encoding
- ✓ Operating System Maintenance
- ✓ Video Editing
- ✓ Computer Drivers

EDUCATIONAL BACKGROUND

TERTIARY	Bachelor of Science in Information and Technology Davao Oriental State University Guang-guang, Dahican, City of Mati Davao Oriental 2018-2022
SECONDARY	Mati National Comprehensive High School (2006-2009)
PRIMARY	Rabat Rocamora Mati Central Elementary School I. (S.Y. 2001-2006)

WORK EXPERIENCE

DSWD Validator
I.net Cafe

KEVIN KHYLE M. HINOJALES
Panamin, Brgy. Matiao
City of Mati, Davao Oriental
+639303343076
kevinkhyle.hinojales@gmail.com

PERSONAL DATA

Date of Birth:	November 06, 1999
Age:	22
Place of Birth:	City of Mati
Sex:	Male
Height:	5'7
Weight:	50 kg
Civil Status:	Single
Nationality:	Filipino

SKILLS AND INTEREST

- ✓ Decision Making
- ✓ Database Management
- ✓ Computer Skills
- ✓ Problem Solving

EDUCATIONAL BACKGROUND

TERTIARY	Bachelor of Science in Information and Technology Davao Oriental State University Guang-guang, Dahican, City of Mati Davao Oriental (2018-2022)
SECONDARY	Sunbeam Christian School of Mati Inc. Mapantad, Brgy. Sainz, Mati City (2011-2015)
PRIMARY	Sunbeam Christian School of Mati Inc. Mapantad, Brgy. Sainz, Mati City (2005-2011)