

Interpreting Predictive Data Analytics Characteristics of Undergraduate Freshmen Relating Choice and Selection of Programs

Richmond Adebaiye¹, Ona Egbue², Theophilus Owusu³

¹Department of Informatics and Engineering Systems University of South Carolina Upstate, USA

²Department of Informatics and Engineering Systems University of South Carolina Upstate, USA

³Graduate Business School Keiser University, Fort Lauderdale, USA

Corresponding Author: Richmond Adebaiye

ABSTRACT

Most freshmen have found a selection of major subjects of study at universities to be challenging due to lack of a specific predictive data analytic criteria upon which they can choose their specialization majors. Researchers have studied the influence of various factors on the selection criteria of study programs and choice of major subjects of study in programs like in business, engineering, health science, applied science, and humanities. However, there is a dearth of research on factors influencing undergraduate freshmen in their choice of study programs and selection of major courses of study in programs like data science and analytics, informatics, information management and systems, and computer science. This study examined the predictive data analytical relationship between the students' characteristics and their learning environmental factors on one hand and the selection criteria of study programs and choice of major subjects of study on the other hand. The research used descriptive survey design (N=211) for data collection and statistical analysis for means, standard deviations and frequencies as well as logistic regression coefficients in answering the research questions. The study found and concluded that age, gender, and ethnicity have no significant influence on selection criteria of study programs and the choice of major subjects of study. The study also found that overall SAT/ACT score, social integration, and self-assessment on mathematical skills, credits attempted, credits completed score, first semester GPA score, self-commitment to complete degree course and self-assessment on degree satisfaction had a statistically significant influence on the choice of study programs and selection of major courses.

Keywords: Predictive data analytics, data analytic criteria, choice of study programs, selection of majors of specialization.

Date of Submission: 20-10-2019

Date of acceptance: 03-11-2019

I. INTRODUCTION

Data analytics has become an immense tool in decision making. Predictive data analytics has been applied to the identification of customer's prospects, retention projections, attrition ratio calculations, detection of fraudulent activities and or default estimates to generate behaviors that impact business objectives (Adebaiye, 2015). Although predictive analytics were applied in higher education for attrition and retention projections and in analyzing student's performances, research has not examined the prospect of effective results emanating from such predictive analysis. According to Bandura, Zimmerman & Martinez-Pons (2012), there is dearth in research on how students "formulate, adjust, and make decisions regarding their academic environments. Other researchers including Maddi (2004) and Tinto (2014) also advocated for theoretical construct self-efficacy, hardiness, and academic integration theory to illustrate the perspective in which undergraduate freshmen choose their majors in a program. The most common use of predictive analytics is in the application of a well-defined set of performance metrics specific to a goal's achievement. Predictive data analytics have equally been valuable in identifying patterns, predict outcomes and guide in the decision-making process (Chau, Ngai, Xiu, 2009; Hersh, 2002). Researchers have studied various factors on the selection criteria of study programs and majors in information security, engineering, health science, applied science and education courses, however, research on the use of predictive data analytics in programs like informatics and computer science is still in its infancy (Tinto, 2014). The selection of major subjects of study at universities is increasingly becoming a challenge among freshmen university students. This could be due to the lack of specific predictive criteria upon which they can rely when choosing and making options among the specialization areas in their degrees of study (Pressler, 2014). The specific predictive data criteria can be developed based on the factors that influence the choice of a given course of study as well as a given major subject of study at university (Tinto, 2014). Allen &

Robbins (2008) noted that the factors that influence the selection criteria of study programs and choice of major subjects of study include; students' demographic characteristics, students' high school academic characteristics, first-semester academic engagement, the university learning environmental factors and the students' degree completion characteristics. Gordon (2007) agrees with Cueso (2005) that; age, gender, and ethnicity are students' demographic characteristics in university learning. He also lists overall SAT/ACT score, math score, high school rank, self-assessment in math ability and self-assessment in scientific skills as students' high school academic characteristics. On the first-semester academic engagement, Ngai, Xiu & Chau (2009) listed factors like the first-semester number of credits attempted, the first-semester number of credits completed, first semester GPA score, first-semester academic self-efficacy and social integration in the first semester. Porter & Umbach (2006) listed academic infrastructural facilities, resource persons like professors and lecturers and a conducive learning environment as the university learning environmental factors that influence the choice of university and program of study. Finally, MacNeill, Campell & Haeksey (2014) cited community membership, honorary members program, self-commitment to complete university degree and self-assessment on degree satisfaction as students' degree completion characteristics that influence on the choice of study program and selection of major subjects of specialization.

1.1 Related work

Literature review on predictive data analytics related to student characteristics and their decision factors in their choice and selection of majors in higher education helped shaped the objectives of this study. Gordon (2007) cited that the selection of major subjects of study at universities has a direct influence on future job opportunities for college graduates. The influence of degree completion outcome characteristics on the choice of study programs could provide self-assessment on degree satisfaction and job opportunities available, and this could strongly be associated with selection criteria of study programs and choice of major subjects of study. On the issues of social isolation affecting undergraduate's choice of study, Chin (2003) concluded that students had "limited opportunities for connecting to the larger community outside of what is provided in individual courses especially when dealing with situations that involve failure/disappointment, or loss of a family member.

Adebiaye (2015) further explained the two types of social isolation in the classroom – "friendlessness and marginalization". This is viewed from the perspective of the "number, frequency and quality of contacts; the longevity or durability of these contacts and the negativism" attributed to the isolation felt by the individual involved. Upah (2016) noted that students have difficulty in selecting an option among majors like; software engineering, computer networking, system analysis & design, programming, cloud computing, software applications, and other related majors. Jenkins & Cho (2012) found the need to develop a predictive data model or criteria based on the significant variables to help in the selection of study programs and choice of major subjects of study in informatics and computing degree programs. Adebiaye (2015) also noted that the predictive data analytic model should consider self-assessment in mathematical skills, basic computer skills, and high school overall SAT/ACT scores which are among the pre-college academic characteristics that will influence the selection of study programs and choice of major subjects of specialization. Hung, Hsu & Rice (2012) also noted that the influence of first semester GPA score, academic self-efficacy, social integration among students in the first semester of study in university and first semester credits completed which are among first-semester academic characteristics should also be considered when making a predictive data model for the choice of study programs and major subjects of study.

Porter & Umbach (2006) cited university learning environmental factors as having an influence on the selection of study programs and the choice of major subjects of study in informatics and computing degree programs. They, therefore, suggest that factors like university learning facilities, university infrastructure, university human resource team, and learning environment should be considered when developing a predictive data analytic model for choice of major subjects of study in informatics and computing degree programs. Baker (2013) suggested that influence of degree completion outcome characteristics like community membership, honorary members programs and commitment to degree completion have an influence on the choice of study programs and selection of major subjects of study. The inclusion of these factors in the predictive data analytic model for choice of major subjects of study is necessary. Norris & Baer (2013) found that there is need to extract data in identifying patterns, trends, and insights to guide actions and decision-making aptly to describe analytics model to enable selection of study programs and choice of major subjects of study in informatics and computing degree programs. A predictive data analytic model should have an ability to collect and integrate data from multiple, varied sources including students, lecturers, university facilities and the job market. Ngai, Xiu & Chau (2009) realized that the predictive data analytic model should help students to be able to formulate, adjust, and make decisions regarding their academic environments in semesters of theoretical construct self-efficacy, academic integration, social integration, and self-satisfaction. Upah (2016) also found that the predictive data analytic model should also provide self-assessment variables like; the level of commitment, subject skills, time management skills and planning, academic life satisfaction, academic self-efficacy, and environmental factors.

The perspectives established by different researchers helped in various ‘desemesterinations’, that the choice of major subjects of study is often viewed as a challenge and the most important decisions an individual needs to make. Many university freshers enter college education with an undecided college major for various reasons. Some researchers have suggested that it is detrimental to offer predictive criteria on which the decision of a major is done. Allen & Robbins (2008) opined that some researchers have posited that attrition increases noticeably among undecided students because of a delay in committing to learning outcomes. Therefore, this study also evaluated the influence of students’ demographic characteristics, socio-economic factors, pre-college entry knowledge, first semester performances, social isolation and degree completion characteristics on their choice and selection of major subjects of study.

II. PURPOSE OF THE STUDY

The objective of this study was to examine the predictive data relationship between the students’ characteristics and their learning environmental factors on one hand and the selection criteria of study programs and choice of major subjects of study on the other hand. The students’ characteristics include; demographic characteristics, social isolation, pre-college academic characteristics, first-semester academic engagement, and degree completion characteristics. The predictive data analytical relationship provided selection criteria that can be used for academic advising on the freshmen to help in ‘desemesterining’ their choice of the major program of study.

2.1 Specific Research Objectives

- i. To establish the influence of demographic characteristics of the first-year student's choice of study programs and selection of major courses of study.
- ii. To ‘desemesterine’ the relationship between the pre-college academic characteristics and selection criteria of study programs and the choice of major subjects of study
- iii. To examine the influence of first-semester academic characteristics on choice of study programs and selection of major subjects of study.
- iv. To examine whether degree completion outcome characteristics have an influence on the choice of study programs and selection of major subjects of study.

2.2 Research questions

The following research questions were derived from the objectives of the study to help in the formulation of the research tools from the web-based questionnaires.

- i. Do the demographic characteristics of the first-year students influence their choice of study programs and selection of major courses of study?
- ii. What is the relationship between the pre-college academic characteristics and selection criteria of study programs and the choice of major subjects of study?
- iii. To what extent does the influence of first-semester academic characteristics on choice of study programs and selection of major subjects of study?
- iv. Do the degree completion outcome characteristics have an influence on the choice of study programs and selection of major subjects of study?

The study is significant because it will help to establish whether demographic characteristics of the first-year students influenced their choice of study programs and selection of major courses of study. In addition, it may also help in formulating a framework for policies that guide the freshmen in their choice of majors and selection of desired programs. Furthermore, the study contributes to the body of knowledge in formulating guiding principles on student’s choice of study programs and selection of majors.

III. METHODS

The research used a descriptive survey design to collect data, coded and organized in software spreadsheets for descriptive and inferential ‘desemesterination’. Creswell (2012) explained that descriptive survey design helps in providing descriptive statistics like means, standard deviations, and frequencies as well as testing the hypotheses on the significant relationships. Such relationships are like those between the students’ academic characteristics and selection criteria of study programs and the choice of major subjects of study.

The study variables included; response variables, explanatory variables, and intervening variables. The response variable of the study was selection criteria of study programs and the choice of major subjects of study. The response variable was ‘desemesterined’ using a 2-point scale with two options namely; Low=1 and High=2. The explanatory or independent variables of the study were categorized as; pre-college academic characteristics, first-semester academic characteristics and degree completion outcome characteristics. The pre-college academic characteristics included overall SAT/ACT score, ACT math score, high school rank, self-assessment

on statistical/mathematical skills and self-assessment on basic computer skills. The first-semester academic characteristics were; first semester credits attempted, first semester credits completed, first semester GPA, first-semester academic self-efficacy and first-semester social integration. The degree completion outcome characteristics were self-commitment to complete degree course, community membership, honorary members program and self-assessment on degree satisfaction. The explanatory variables were measured using a 5-point Likert scale with options being: Very Low=1, Low=2, Moderate=3, High=4, and Very High=5. The intervening variables included; personal demographic attributes of the study participants such as age, gender, ethnicity, and social-economic status.

3.3 Data collection and analysis

The study targeted undergraduate freshmen studying business/data analytics, informatics, and computer science. In addition, the study used simple random sampling in which each freshman had an equal chance to be selected to form a sample of the study. The questionnaires were randomly distributed to freshmen and had most responses in semesters of 5-point scale ratings: very low, low, moderate, high, and very high. The 5-point scale ratings were to ensure reliability, validity, and precision features for statistical analysis. The socio-demographic data were analyzed using descriptive statistics, frequency tables, and graphical methods. The data concerning research objectives' study variables were analyzed using descriptive statistics, Student's t-test and logistic regression analysis. The descriptive statistics employed mean (μ) and standard deviation (σ) values in the analysis while logistic regression analysis, the regression coefficients (B), Z and P-values were used to address the research questions.

IV. RESULTS OF THE STUDY

This section presents the results of data analysis on the socio-demographic factors of the study participants and the analysis concerning research questions.

Demographic factors

Table 1: Frequency distribution and descriptive statistics

Variable	Attribute	Frequency	Percent	Mean	Standard deviation
Age in years	16	25	11.8	17.78	1.042
	17	57	27.0		
	18	78	37.0		
	19	41	19.4		
	20+	10	4.7		
	Total		211		
Gender	Male	111	52.6	1.52	0.580
	Female	91	43.1		
	Partners	9	4.3		
	Total	211	100.0		

The results in Table 1 indicate that the plurality of the respondents was 18 years old as at the time of the research with a relative frequency of 37.0%, followed by those who were 17 years and 19 years old with relative frequencies of 27.0% and 19.4% respectively. Those aged 16 years and more than 20 years old were minor groups with relative frequencies of 11.8% and 4.7% respectively. The results also indicated that the mean age of respondents was 17.78 years with a standard deviation of 1.042 ($\mu=17.78$ and $\sigma=1.042$). The age of the freshmen students displayed a normal curve as shown in the histogram below.

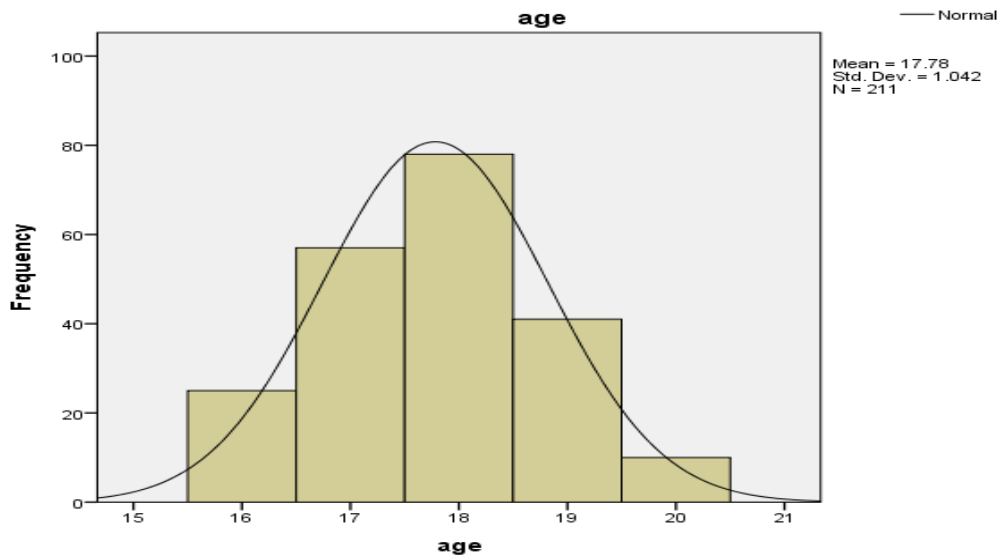


Figure 1: Histogram of age distribution

Concerning the gender, the results in Table 1 show that 52.6% of the study participants were males while 43.1% were females. In addition, 4.3% of the study participants identified themselves as partners. These results show more males than females among the freshmen and college-first-semester students studying data analytics, informatics or computer science. Therefore, results indicate that gender has an effect on the choice of study and selections of majors.

Table 2: Frequency distribution table of response variables

Variable	Attribute	Frequency	Percent
Selection criteria of study programs	Low	64	30.3
	High	147	69.7
	Total	211	100.0
Choice of major subjects of study	Low	37	17.54
	High	174	82.46
	Total	211	100.0

Those who agreed that there were selection criteria (predictive analytical criteria) represented in helping students choose their programs were 147 out of 211 with a relative frequency of 69.7% while 64 disagreed that there was a selection criterion of study programs. On the other hand, 82.46% agreed while 17.54% disagreed that universities had predictive data analytical criteria provided to students in enabling choices.

4.1 Analysis of Study objectives

Table 3: Frequency distribution table of independent variables

Variable	Factor	Responses					Total
		Very Low	Low	Moderate	High	Very High	
The pre-college academic characteristics	Overall SAT/ACT score	0	15.2	17.1	34.1	33.6	100.0
	SAT/ACT math score	0	0	32.2	42.7	25.1	100.0
	High school rank	0	5.2	10.4	47.4	37.0	100.0
	Self-assessment on mathematical skills	7.1	12.3	23.7	22.7	34.1	100.0
	Self-assessment on basic computer skills	9.0	8.1	43.1	24.6	15.2	100.0
The first semester academic characteristics	Credits attempted	5.2	9.0	18.5	55.9	11.4	100.0
	Credits completed	0.5	18.0	15.6	37.4	28.4	100.0
	first semester GPA	5.7	9.5	9.0	55.9	19.9	100.0
	1 st semester academic self-efficacy	6.6	12.8	64.9	10.0	5.7	100.0

	1 st semester social integration	2.8	47.4	19.9	14.2	15.6	100.0
The degree completion outcome characteristics	Self-commitment to complete degree course	3.8	5.7	25.1	39.3	26.1	100.0
	Community membership	13.3	53.6	10.9	19.4	2.8	100.0
	Honorary members program	5.7	40.3	38.4	8.5	7.1	100.0
	Self-assessment on degree satisfaction	7.6	10.9	11.4	53.6	18.0	100.0

The results in table 3 show that the majority of study respondents rated the influence of overall SAT/ACT score on the choice of study programs and selection of major courses of study as either high or very high with relative frequencies of 34.1% and 33.6% respectively. The results also indicate that 42.7% and 25.1% of respondents rated the influence of SAT/ACT math scores on the choice of study programs and selection of major courses as high and very high respectively. Concerning high school rank, 47.4% and 37.0% of respondents rated the influence as either high or very high. The study also found that 22.7% and 34.1% of the respondent rated the influence of self-assessment on mathematical skills on the choice of study programs and selection of major courses as high and very high respectively. The results also displayed that the majority of the respondents rated the influence of self-assessment on basic computer skills on the choice of study programs and selection of major courses as moderate with a relative frequency of 43.1%. Therefore, overall SAT/ACT score, SAT/ACT math score, high school rank and self-assessment on mathematical skills have either high or very high ratings on influence on the choice of study programs and selection of major courses.

The study also sought the influence of the first-semester academic characteristics, the results in table 3 show that majority of respondents rated the influence of credits attempted, credits completed and first semester GPA on the choice of study programs and selection of major courses as either high or very high. The relative frequencies are: Credits attempted (H=55.9% & VH=11.4%), credits completed (H=37.4% & VH=28.4%), first semester GPA (H=55.9% & VH=19.9%). The results, however, show that the first-semester academic self-efficacy was rated to have a moderate influence on the choice of study programs and selection of major courses with a relative frequency of 64.9%. Concerning the first semester of social integration, 47.4% of the respondents rated their influence on the choice of study programs and selection of major courses as low. The study also sought to determine the influence of the degree completion outcome characteristics on choice of study programs and selection of major courses. This shows a relative result spanning self-commitment to complete degree course (H=39.3% & VH=26.1%) and self-assessment on degree satisfaction (H=53.6% & VH=18.0%) had an either high or very high rating of influence on the choice of study programs and selection of major courses. The results also indicate that the majority of respondents rated the influence of community membership (L=53.6%) and honorary members program with peers on campus (L=40.3%) on the choice of study programs and selection of major courses as low. Therefore, the study found that self-commitment to complete degree courses and self-assessment on degree satisfaction had an either high or very high influence on the choice of study programs and selection of major courses.

Table 4: Logistic regression results on selection criteria of study programs and choice of major subjects of study

Variables in the Equation	B	S.E.	Wald	df	Sig.	Exp(B)
Age	.081	.007	.047	1	.329	.080
Gender	.172	.039	.595	1	.875	1.011
Ethnicity	.314	.127	.445	1	.310	.714
Overall ACT score	.647	.315	4.216	1	.040	1.911
ACT math score	.548	.171	4.124	1	.042	.706
High school rank	.284	.159	3.178	1	.075	.753
Self-assessment on mathematical skills	.432	.191	5.087	1	.024	.649
Self-assessment on basic computer skills	.075	.248	.091	1	.763	.928
Credits attempted	.827	.263	12.388	1	.000	.396
Credits completed	.660	.264	6.231	1	.013	.517
first semester GPA	.782	.287	7.411	1	.006	.458
1 st semester academic self-efficacy	.020	.207	.009	1	.923	.980
1 st semester social integration	.072	.234	.095	1	.758	1.075
Self-commitment to complete degree course	.753	.227	11.046	1	.001	.471
Community membership	.108	.217	.246	1	.620	1.114
Honorary members program	-.139	.202	0.738	1	.130	.645

Self-assessment on degree satisfaction	1.547	2.228	4.164	1	.041	94.321
--	-------	-------	-------	---	------	--------

The study used logistic regression to establish the relationship and influence of demographic characteristics, the pre-college academic characteristics, first-semester academic characteristics, and degree completion outcome characteristics on selection criteria of study programs and choice of major subjects of study. The results in table 4 show that overall SAT/ACT ($B=0.647$ & $p=0.040$), SAT/ACT math score ($B=0.348$ & $p=0.042$) and self-assessment on mathematical skills ($B=0.432$ & $p=0.024$) were significant pre-college academic characteristics that influence selection criteria of study programs and choice of major subjects of study. Concerning the first semester academic characteristics, the research found that; credits attempted ($B=0.827$ & $p=0.000$), credits completed score ($B=0.660$ & $p=0.013$) and first semester GPA ($B=0.782$ & $p=0.006$) influence selection criteria of study programs and choice of major subjects of study. Finally, the study realized that self-commitment to complete degree course ($B=0.753$ & $p=0.001$) and self-assessment on degree satisfaction ($B=1.547$ & $p=0.041$) were significant degree completion outcome characteristics that influence selection criteria of study programs and choice of major subjects of study.

The study also realized that age, gender, ethnicity, high school rank, self-assessment on basic computer skills, first-semester academic self-efficacy, first-semester social integration, community membership, and honorary members program on campus with peers had p-values that were less than 0.05. Therefore, based on a cut-off value of 0.05 for p-values, these factors had no significant influence on the selection criteria of study programs and the choice of major subjects of study.

The logistic regression coefficients in table 4 show that; if the overall SAT/ACT score is increased by one level, the influence of selection criteria of study programs and choice of major subjects will increase by 64.7% ($B=0.647$) while when the SAT/ACT math score is increased by one level, the influence of selection criteria of study programs and choice of major subjects will increase by 34.8% ($B=0.348$). Similarly, results show that when self-assessment on mathematical skills improves by one level, the influence of selection criteria of study programs and choice of major subjects will increase by 43.2% ($B=0.432$) provided all pre-college academic characteristics are held constant.

The logistic results also portray that if first semester credits attempted are increased by one, then the influence of selection criteria of study programs and choice of major subjects will increase by 82.7% ($B=0.827$). Also, if credits completed score improves by one level, the influence of selection criteria of study programs and choice of major subjects will increase by 66.0% ($B=0.660$) and finally if the first semester GPA increases by one, the influence of selection criteria of study programs and choice of major subjects will increase by 78.2% ($B=0.782$) provided all other first semester academic characteristics are held constant. Finally, the study realized that; if self-commitment to complete degree course is increased by one level, then the influence of selection criteria of study programs and choice of major subjects of study will increase by 75.3% ($B=0.753$) while if self-assessment on degree satisfaction is increased by one level, the influence on selection criteria of study programs and choice of major subjects of study will increase by 154.7% ($B=1.547$) provided other degree completion outcome characteristics are held constant.

V. DISCUSSION

Based on the results the study realized the following findings and conclusions were based on the four research objectives.

5.1 The demographic characteristics of the first-year students

The study found that the plurality of the respondents was 18 years old at the time of the research with a relative frequency of 37.0%, followed by those who were 17 years and 19 years old with relative frequencies of 27.0% and 19.4% respectively. The study also found that the mean age of university freshmen and first-semester college students was 17.78 years with a standard deviation of 1.042. The study also found that there were more males than females among university freshmen and first-semester college students, while a paltry 4.3% of students identified themselves as partners (or do not want to be identified as male or female or transgender). The study through logistic regression found that age and gender had no significant regression coefficients (p -values > 0.05). Therefore, the study concludes that demographic characteristics have no statistically significant influence on the selection criteria of study programs and the choice of major subjects of study.

5.2 The relationship between the pre-college academic characteristics and selection criteria of study programs and the choice of major subjects of study

The study found that overall SAT/ACT score, SAT/ACT math score, high school rank and self-assessment on mathematical skills have an either high or very high rating on influence on the choice of study programs and selection of major courses. The research through logistic regression concludes that; overall SAT/ACT, SAT/ACT math score and self-assessment on mathematical skills were significant pre-college academic characteristics that influence selection criteria of study programs and choice of major subjects of

study. The research also found that; increase in the overall SAT/ACT score by one level will increase the influence of selection criteria of study programs and choice of major subjects by 64.7%. While when the SAT/ACT math score is increased by one level the influence of selection criteria of study programs and choice of major subjects will increase by 34.8%. Similarly, when self-assessment on mathematical skills improves by one level the influence of selection criteria of study programs and choice of major subjects will increase by 43.2%.

5.3 The influence of first-semester academic characteristics on choice of study programs and selection of major subjects of study

From the descriptive statistics analysis, the study found that course credits attempted, credits completed, and first semester GPA scores had either high or very high influence on the choice of study programs and selection of major courses. The logistic regression analysis found that credits attempted, credits completed score and first semester GPA score had a significant influence on the selection criteria of study programs and choice of major subjects of study. The study concludes: When first semester credits attempted are increased by one, then the influence of selection criteria of study programs and choice of major subjects will increase by 82.7%. Similarly, if credits completed score improves by one, level the influence of selection criteria of study programs and choice of major subjects will increase by 66.0% and if the first semester GPA increases by one, the influence of selection criteria of study programs and choice of major subjects will increase by 78.2%.

5.4 The degree completion outcome characteristics influence the choice of study programs and selection of major subjects of study

The study found that self-commitment to complete degree course and self-assessment on degree satisfaction had high and very high ratings on the influence on the choice of study programs and selection of major courses. From the logistic regression analysis, the study found that if self-commitment to complete degree course is increased by one level, then the influence of selection criteria of study programs and choice of major subjects of study will increase by 75.3% while if self-assessment on degree satisfaction is increased by one level, the influence on selection criteria of study programs and choice of major subjects of study will increase by 154.7%.

VI. CONCLUSIONS AND RECOMMENDATIONS

Institutions of higher education invest significant resources towards helping students achieve optimal decisions in their selection of programs of study and courses. However, despite the prevalence of predictive data analytics for guiding decision-making processes, this field has limited application for program and course choice in higher education. This paper uses predictive data analytics to examine the factors, including pre-college and first-semester academic characteristics as well as demographic and degree completion characteristics, that influence students' choice of study programs or major courses particularly in informatics and computer science. Based on logistic regression, the study found that several pre-college academic characteristics (overall SAT/ACT score, SAT/ACT math score, and self-assessment on math skills) and first-semester academic characteristics (credits attempted, credits completed score and first semester GPA) significantly influence the choice of study programs and the selection of major courses. In addition, logistic regression results revealed that degree completion outcome characteristics (self-commitment to complete degree and self-assessment on degree satisfaction) have a significant influence on the choice of study programs. However, regression results showed that the influence of the three demographic characteristics assessed including age, gender, and ethnicity on study program choice and major course selection was not statistically significant.

The results of this study have implications for business/data analytics/informatics and computer science programs but also for colleges and universities in general. By providing insight into factors that attract students to business analytics, data science, informatics, and computer science programs of study, this research created recruiting, educational and research implications. In order for students to make informed decisions about the choice of a particular program or major courses, it is important that they are made aware of expectations, opportunities and challenges of each program and that colleges and universities consider the effect of both institutional and personal factors on the selection process.

The results of this study suggest that universities offering business courses especially data analytics, data science, informatics and computer science programs should guide both existing and potential students in the selection process by providing them with information on the specific knowledge and skills needed for success in selecting and determining their majors. This includes pre-college academic requirements and early college (particularly first semester) academic performance. Since the first semester GPA and credits attempted and completed were shown in this study to significantly influence the choice process, the types of courses students are exposed to during this period is critical to their decision-making and success in a program. For

instance, depending on the courses taken during the first semester, students may decide to major in business or data analytics, informatics or computer science without taking a course in either discipline or without taking courses that are good predictors for performance in these disciplines. This could have implications for successful completion of the program. The degree completion outcome characteristics that were shown to be significant were based on degree satisfaction and self-assessment of ability to complete the degree which is likely influenced by other personal and behavioral factors. Furthermore, personal, behavioral and other social factors are likely to affect how students assess their first-semester academic performance and their math skills. Finally, it is necessary to conduct future research to further examine how underlying social, cultural and psychological factors related to students' perception of self-efficacy, math skills, introductory courses, and the requirements of business or data analytics, data science, informatics, and computer science professions affect their selection process.

REFERENCES

- [1]. Adebaiye, R. S., (2015). "Interpreting Crisis of Attrition and Social Isolation in Undergraduate Degree Programs using Asynchronous Learning System". *International Journal of Advanced Scientific Research & Development (IJASRD)*, 02 (01/I), pp. 01 – 21.
- [2]. Allen, J., & Robbins, S. B. (2008). Prediction of college major persistence based on vocational interests, academic preparation, and first-year academic performance. *Research in Higher Education*, 49(1), 62-79.
- [3]. Baker, R. S. J. d. (2013). Learning, schooling, and Data Analytics. In M. Murphy, S. Redding, & J. Twyman (Eds.), *Handbook on innovations in learning for states, districts, and schools* (pp.179–190). Philadelphia, PA: Center on Innovations in Learning.
- [4]. Creswell, J.W. (2012). *Qualitative inquiry and research design: Choosing among five approaches*. Thousand Oaks, CA: Sage.
- [5]. Ferrare, J. J., & Lee, Y. G. (2014). Should we still be talking about leaving? A comparative examination of social inequality in undergraduate patterns of switching majors. WCER Working Paper, 2014-5. Retrieved from <http://files.eric.ed.gov/fulltext/ED556490.pdf>
- [6]. Gordon, V. N. (2007). *The undecided college student: An academic and career advising challenge*. Springfield, IL: Charles C Thomas Publisher.
- [7]. Hung, J. L., Hsu, Y. C., & Rice, K. (2012). Integrating data mining in program evaluation of K-12 online education. *Journal of Educational Technology & Society*, 15(3), 27-41.
- [8]. Jenkins, D., and Cho, S.-W. (2012). *Get With the Program: Accelerating Community College Students' Entry into and Completion of Programs of Study* (CCRC Working Paper No. 32). Community College Research Center. New York, NY: Teachers College, Columbia University.
- [9]. MacNeill, S., Campbell, L. M., & Hawksey, M. (2014). Analytics for education. *Journal of Interactive Media in Education*. Retrieved from: <http://jime.open.ac.uk/jime/article/viewArticle/2014-07/html>
- [10]. Ngai, E. W. T., Xiu, L., and Chau, D. C. K. (2009). Application of data mining techniques in customer relationship management: A literature relationship and classification. *Expert Systems with Applications*, 36(2), 2,529-2,602.
- [11]. Norris, D. M., & Baer, L. L. (2013). Building organizational capacity for analytics. *EDUCAUSE Review Online*, 1-58. Retrieved from <https://net.educause.edu/ir/library/pdf/PUB9012.pdf>
- [12]. Porter, S. R., & Umbach, P. D. (2006). College major choice: An analysis of person-environment fit. *Research in Higher Education*, 47(4), 429-449.
- [13]. Pressler, E. J. (2014). Logging in to learning analytics. *Current Issues in Emerging eLearning*, 1(1), 6.
- [14]. Tinto, V. (2014). Tinto's South Africa lectures. *Journal of Student Affairs in Africa*, 2(2).
- [15]. Upah, S.C. (2016) *The impact of predictive data analytics based advising on the selection and change of major among first-year, first-semester students in engineering*. University Ph.D. dissertation. Iowa State University. Ames, Iowa.