**CHAPTER-1**

**INTRODUCTION**

Whatever the decisions that are made at the strategic level which is handled by the managing board of any higher educational institutions affect the polices, strategies and actions.

The disconnection between the authoritative stakeholders and the effective computational algorithms lead to the following discomforts, they are:

1. Longer decision process.
2. The whole data is not involved inside the decision-making process.
3. A small predictable impact on the academics is caused.
   1. **Why does the disengagement occur?**

Higher education agencies and educational policy makers, students and their subordinates are chronically disengaged due to the following few reasons:

1. Lot of barriers that includes technological conditions, non-flexible governance, continuous changes in the governing structure.

With regards to the technological conditions, there are significant obstacles in the face-to-face educational model of the modern world. Taking an example, the administrative and academic knowledge are housed in separate silos making working formats substantially varying.

1. The observations in our case study point towards the transactional processes such as the pupils’ attendance record or graduate documents, to point a few are still done manually in paper notes.

Nonetheless the distance-educational model arenas offer more resources and flexibility towards the information. The consignment of rationalized and standardized information is inborn, as computers are eminently ideal and empirical for the concerned work.

1. The decision of the directors and managers does not stand strong when the data stored in the silos is taken as the base data. To provide an appropriate presentation of valuable educational data, whether systematically stored or manually stored, these must be analyzed thoroughly to support complex processes.
   1. **Decision classification at higher educational institutions**

Higher educational institutes come under the tertiary sector of organizations. Governance of finances, activities and personnel by retaining the autonomy on decisions are the primary responsibilities of these organizations.

There are three types governance observed at the HEIs which influences the operational and managerial behavior: Academic, Bureaucratic and Corporate

1. **Academic**

The members of the faculty work together to retain authority decision-making powers in areas such s teaching, academics, administration, curriculum etc.

1. **Bureaucratic**

University retains layers of hierarchy with work divisions characterized by policies, upper-level departments and clear orders.

1. **Corporate**

The universities around follow the practice of enterprise to highlight customer needs and market competition. Often the universities prefer and implement academic and bureaucratic styles. Among the HEIs, there are particular differences in terms of mission and management strategies, for an example, private universities are more leaned towards the market and actions whereas the public universities take students and alumni as their cornerstone.

**Decision Classification**

To divide the work vertically pertaining to the decisions’ a hierarchic structure is followed which resembles a pyramid structure.

Strategic

Tactic

Operational

Figure 3: Decision structure levels at HEIs

1. **Strategic**

This is the uppermost level which defines the policies and strategies that are needed for the organization integrating the primary goals and actions into a consistent world. The governing board, rector and deans constitute the managing positions. The decisions on critical factors in strategic planning is made by them.

1. **Tactic**

This level identifies and executes the detailed plans made at the strategic level. Usually, deans work with the heads of the departments to achieve the interests and goals. The intermediate directors help in coordinating the resources. Once the strategic plan is accepted then the tactic level of planning comes into play.

1. **Operational**

This is the lowest level; it is in charge of the day to day processes. Particular tasks and activities are performed to support operations in the institutions. This level holds the majority of IT required by the HEIs.

* 1. **How does Machine Learning help overcome the disengagement issue?**

The Machine Learning is an emerging field under artificial intelligence that uses various algorithms and analyzes the provided information and provides a better understanding of the data contained in a particular content. In this project we focus upon supporting the decision-making process at the strategic level, considering the deans ‘foremost mission to strengthen the graduation rates.

In this project, we take three supervised classification algorithms to predict the graduation rates from real time data about the undergraduate engineering students.

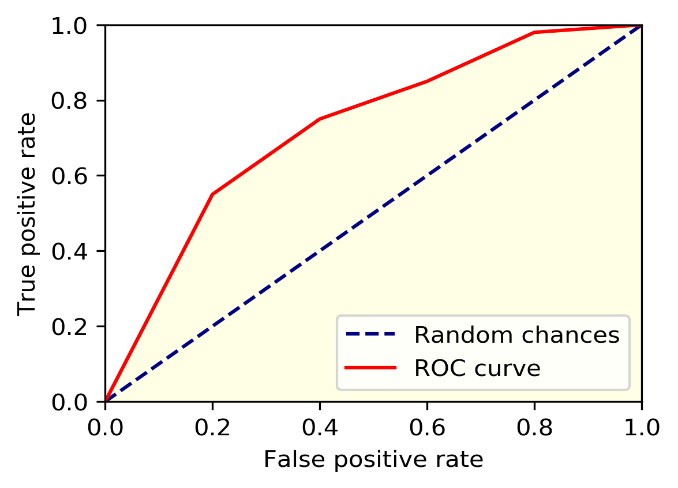
The study of the receiver operating characteristic (ROC) curve and accuracy is carried out as efficiency measure to compare and test decision tree, logistic regression, and random forest where the latter produces the best results.

Figure 1.2 Representation of the ROC curve

**CHAPTER-2**

**LITERATURE SURVEY**

**2.1** **The effect of blended learning on student performance at course-level in higher education. [1]**

This paper analysis the impact of blended learning on the academic achievements of higher education students. Included the discipline and instructors end of evaluation method as variables. Although being considered as a small effect according to standard criteria (Cohen’s d), the weighted mean effect is significant, which supports the perspective that BL can result in better learning outcomes for higher education students Blended learning demonstrates a small summary effect (g\* = 0.385, p&lt;0.001) compared to the traditional teaching methods. The present research adds value in this field with a meta-analysis considering the moderating effect of disciplines and assessment methods. Therefore, the results encourage the implementation of BL in STEM-disciplines, with a weighted mean effect size of g+ = 0.496 when contrasted with face-to-face learning conditions. The study confirms the effect of BL on student performance in a higher education setting only. Thus, the finding prompts instructors in non-STEM disciplines to be more attentive to facilitating constructive and critical online discourses if BL is to bring about high learning quality compared to other modes of delivery.

**2.2 A machine learning framework to identify students at risk of adverse academic outcomes [2]**

This paper analysis the school districts that have developed successful intervention programs to help student graduate high school on time. The work described in this paper was done in collaboration with two school districts in the US (with combined enrollment of around 200,000 students) and is aimed at giving them (as well as other schools) proactive tools that are designed for their needs, and to help them identify and prioritize students who are at risk of adverse academic outcomes. Although the work in this paper is limited to predicting students who are likely to not finish high school on time, we believe that the framework (problem formulation, feature extraction process, classifiers, and evaluation criteria) applies and generalizes to other adverse academic outcomes as well, such as not applying to college, or undermatching The perineal challenge faced by schools today is to improve the student graduation rates. The success of individual intervention programs depends on school’s ability to accurately prioritize students who need help. Could help the schools and institution to predict the pass rate of the students that are admitted. The work in this paper is limited to predicting students who are likely to nit finish high school on time, this framework generalizes to other adverse academic outcomes as well

.**2.3 Fairness in examination timetabling: Student preferences and extended formulations** **[3]**

This paper talks about the variation of timetabling problem. The common characteristics between all problems is the fact that the definite and data sets used, all originate from actual educational institutions, particularly universities, including specific examination criteria. Our survey of student views found that over half of them were unhappy with their examination timetables. Furthermore, about 30% of respondents even believed that their examination timetable negatively affected their academic achievement. We have no evidence that the timetables actually did affect student performance, but the perception is important; especially with Universities competing for students. Therefore, this work intends to contribute to generating examination timetables that match student preferences and enhance their satisfaction. In particular, we have proposed and studied methods to improve fairness amongst students in the timetables they receive. A crucial contribution of this paper is to introduce the novel concept of `fairness within a cohort of students; this complement and widens the concept of fairness within the entire student bodies. This work utilizes the results of an extensive survey seeking student preference with regard to their individual examination. This work intends to contribute to generating examination timetables that match student preferences and enhance their satisfaction. This work intends to only generalizing the method of timetabling process not taking other factor of the students into consideration.

.**2.4 Private colleges teachers evaluation system based on support vector machine (SVM) [4]**

This paper brings to light that private colleges can gain a sustainable development through a good social influence; the key is to use the quality of personal training. Using structural risk minimization principle, SVM for cases of statistical learning of small sample problem have very good classification ability and generalization ability [3]. This article will support vector machine (SVM) theory is introduced into the private colleges of the evaluation of teachers, teaching evaluation of teachers. The success of the support vector machine (SVM) to a large extent depends on the Kernel function technology, Support vector machine (SVM) model selection problem is given a kernel function, by adjusting the parameters of the nuclear and penalty factor. In the teaching evaluation, the evaluation process is generally determined according to the actual situation of school teachers evaluation index, and then according to the index to evaluate teachers; data collection. In this paper, according to the current widespread teacher’s evaluation index established evaluation index system, selection of a private colleges and universities over the years according to the teachers, students, and unfairness feedback that evaluate the most reasonable 208 samples as training sample and test data. Each set of data has 12 evaluation indexes, namely feature vector of each sample data a total of 12 properties. The principle of SVM evolved the optimal separating hyperplane. Support vector machine (SVM) method for small sample data has good classification ability and generalization ability and learning, better solve the small sample, nonlinear, high dimension and local minimum point and so on practical problems.

**2.5 Student engagement predictions in an e-Learning system and their impact on student course assessment scores [5]**

Several challenges are associated with e-learning systems, the most significant of which is the lack of student motivation in various course activities and for various course materials. In this study, we used machine learning (ML) algorithms to identify low engagement students in a social science course at the Open University (OU) to assess the effect of engagement on student performance. This paper talks about the web-based learning that has increased in today’s world. It has become a common place in education and can take many forms, from massive open online courses (MOOCs) to virtual learning environment (VLE) and learning management system. Due to absence of face to face meetings, web-based system faces some challenges that need to be addressed. Could predict the rate of low engagement students on the e-learning platform. This framework was costly and could not be affordable by a common man.

**CHAPTER-3**

**EXISTING SYSTEM**

Traditionally the reports of the students and the other required stats were given by face to face interaction between the students, teachers and the management as well as the parents of the students.

The process the occurred in the existing system is as follows:

* The data of required comes from a face to face model. That is, every data that is required is taken from the students by their performances directly, and this is then produced in front of the management for further decisions.
* The algorithms used differ due to the higher ratios of data set used
* When it comes to the stakeholders in the present model, they are just the students and teachers
* The decisions are supported strategically when the right information is given to the high chain management.

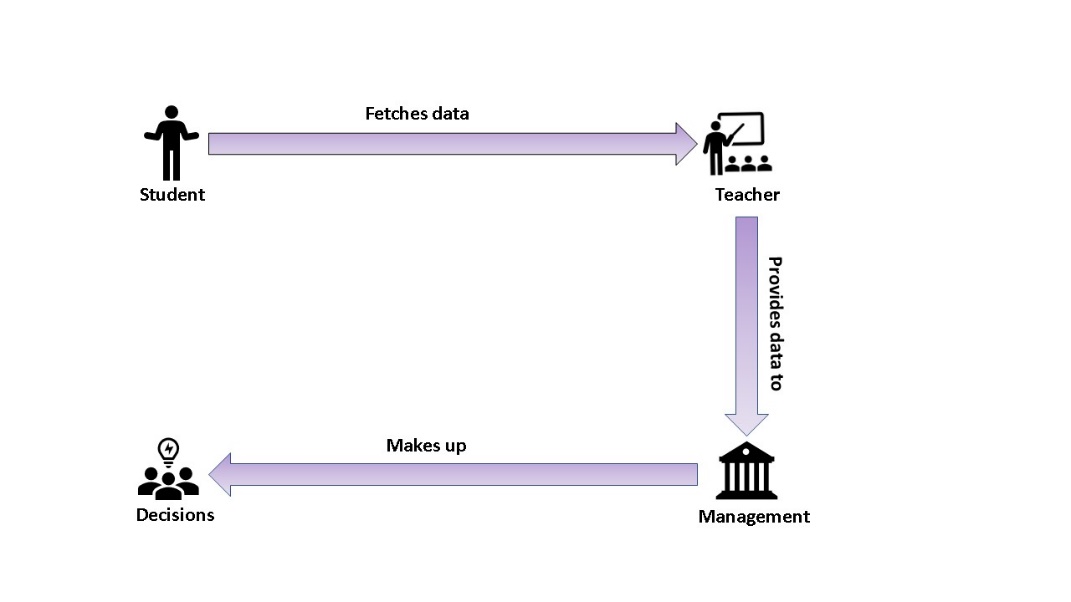
**3.1 Existing System Model**

|  |  |  |  |
| --- | --- | --- | --- |
| **HEI** | **Goals** | **Stakeholders** | **Algorithms used** |
| University of Liege  Sep 2017 | Predict freshman’s failure. Used data from school records and environmental factors | Students and advisors | RF,LR,ANN |
| University of Alagoas  Aug 2017 | To predict students’ failure rates in inductory courses to give educator’s questions | Teachers and tutors | DT,SVM,ANN,NB |
| Babcock University  May 2015 | To predict students’ performance and recommend necessary action. | Advisors and tutors. | DT,RFF,ANN |
| MIT 2014 | To predict and intervene in final course scores using process level information**.** | Teachers and tutors. | LIR,LR,ANN,SVM |
| Amrita school of engineering Jan 2017 | To predict the placement performance of students | Students and teachers | DT,LR |

**Fig 3.1 Table showing existing systems**

The above table shows all the existing systems that are being used in the prominent universities around the globe. When it comes to the stake holders it is mostly the students, advisors and the

teachers.

****

**Fig 3.1 Table showing existing systems**

As shown in the above image the teachers or the advisors take in data from the students regarding their academic and co-curricular activities, then these data are then given to the management board who come out with decisions using some of their strategies.

**3.2 Summarized Procedure**

Main contributions can be summarized as follows:

* Data comes from a face-to-face educational model.
* Due to more features are included, algorithms, architecture differ.
* The stakeholders are students and teachers from HEIs which have particular visualization results
* Strategic decisions are supported when the right information is given to the high chain management

**3.3 Disadvantages of Existing System**

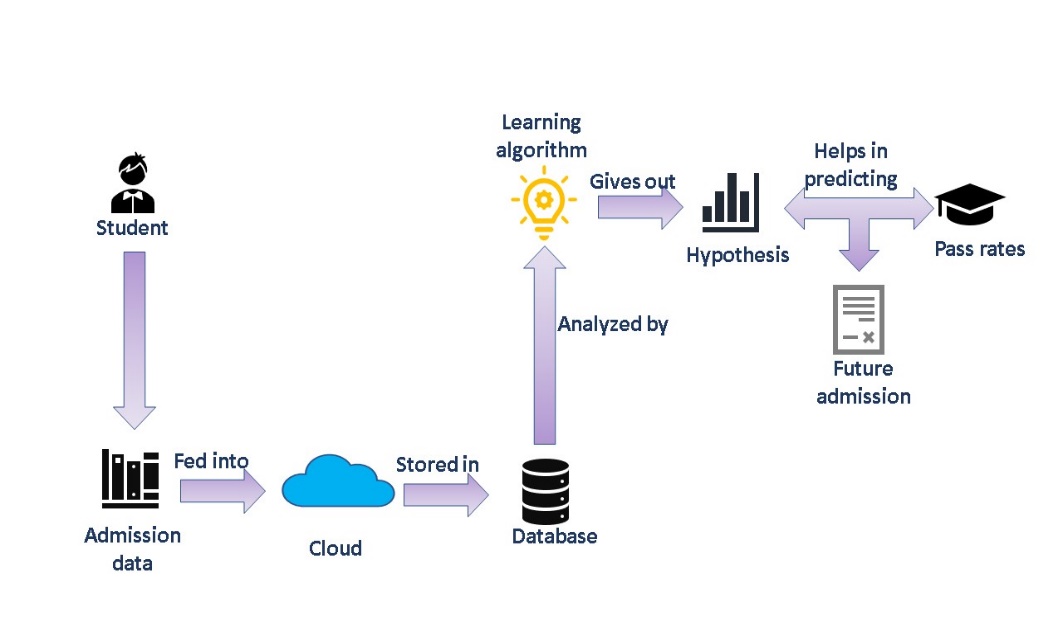
* The stakeholders are only the teachers and students not considering the deans and directors of the HEIs.

**CHAPTER 4**

**PROPOSED SYSTEM**

We propose a better and efficient model for the problem that is faced by the higher education institutes. The efficiency is brought about by the machine learning framework that is being used in the system.

**4.1 Proposed System Model**



**Fig 4.1 Proposed System**

As shown in Fig.4.1, the students data is being fed into the server and that data from the server is then stored in a database in the form of tuples and then this tabulated data is analyzed by the learning algorithm that gives us an hypothesis and this hypothesis helps us in predicting the pass rates and future admissions.

**4.2 Overview**

* In-depth analysis of the number of features tested, or data normalization, will be presented.
* Include analysis of effectiveness metrics such as F-Measure or specificity as well as the comparison with other classification algorithms.
* Include socio-demographic, socio-economical information, involvement in other co-curricular activities information about the students when analyzing the variables that might affect the graduation rates at Higher Educational Institutions.

**4.3 Methodology**

**4.3.1 Data Collection and preparation**

The experiment was conducted with real data from a public university in Colombia. The data source contains information from 6100 engineer students. We analyzed students from five different engineer programs such as industrial, cadastral, systems, electrical and electronical engineering, enrolled during the ten years period of 2004-2014. This number of students is assumed after disregarding cases of missing data such as students who leave during the first semester and students who enter the university after 2009. Students who enter after 2009 will not graduate before 2014 because engineering careers take ten semesters to graduate and therefore will not address the supervised algorithms needs. More than 55200 records were available to analyze. As our case study is a public university, data policies are strict. Although our research was restricted by their data protection policies, and we lack information about students' gender or age among other socio-demographic data, for the most part we use students' academic records to held the graduation rates prediction. We believe the inclusion of sociodemographic and socio-economic data would be worth to analyze in the future. However, in this study the academic information obtained is efficient to analyze the insightful outcomes. Once the classification objective was set (i.e., graduated and not graduated students) and data was acquired, we conducted the following steps to build every algorithm model:

i. Using the scaling method, we transformed data by giving them values from a range [0, 1] 0 as a minimum and a

maximum of 1.

ii. Set initial hyperparameters for each algorithm.

iii. Using a stratified sampling technique, split the dataset into two subsets 70% for training and 30% for prediction,

to keep the data distribution. The sampling method alleviate the effect of class imbalance problem as one of the most

employed method.

iv. From the training subset in step three, we use 5-fold cross validation technique to tune the hyperparameters in

each algorithm.

v. We execute the algorithms with the initial settings. In each k-fold we save the accuracy obtained as well as the

values entered in each hyper parameter, to adjust them in each run and encounter the most suitable values for them until the accuracy reached the expectations.

vi. Finally the algorithms are executed to train the whole training set using the best values obtained for the hyperparameters in each algorithm. Hyperparameters set in each algorithm as well as the architecture and contextualization of each algorithm is exposed in the next subsection.

**4.3.2 Machine learning algorithms**

1. **DECISION TREE(DT)**

Due to its effectiveness of simplicity for being understood and for its interpretation it is a highly used classifier. The prerequisites for this algorithm are data preparation, case study of the material, studying the classification. DT falls under the supervised learning algorithm. Even though there are many specific DT algorithms, the algorithm that has been used in this project is the one that is developed by Quinlan. C5.0 is significantly faster and efficient than ID3 which gives the tree more accuracy.

Furthermore, misclassification types and weighting of different attributes is allowed by this algorithm. While constructing DT it is important to find the best splitting point measure. Gini index is used as a split measure for choosing the most appropriate splitting attribute for each node. The split function has the form:

**Ig (p) = 1-∑ji=1 p2i**

Gini index is mathematically shown above; with J classes suppose 1 2 {1,2,3,…J) and pi is the fraction of items labeled with class ‘i’ in the set.

1. **Random Forest (RF)**
2. **Logistic Regression (LR)**

**Chapters that are needed to be added!**

**CHAPTER 5**

**SYSTEM ARCHITECTURE**

**CHAPTER 6**

**SYSTEM DESIGN AND ANALYSIS**

\

**CHAPTER 7**

**SYSTEM TESTING**

**CHAPTER 8**

**SYSTEM REQUIREMENT SPECIFICATION**

**CHAPTER 9**

**SYSTEM IMPLEMENTATION**

**SNAPSHOTS**