KALBOARD 360 STUDENT PERFORMANCE PREDICTION AND EVALUATION BY MACHINE LEARNING

By Bhanu Sreekar Reddy Karumuri

OVERVIEW

- Prior Research
- Research Hypothesis
- Dataset Description
- Methodology
- Results
- Challenges faced
- Acknowledgement

PRIOR RESEARCH

- Most of the previous research was focused on understanding the importance of a particular feature in predicting the academic performance of students.
- One-third of the papers considered cumulative grade point average(CGPA) as the best predictor for future academic performance.
- Other papers considered finding the correlation between academic performance and family income and found a strong positive correlation between them.
- Other papers considered parenting styles, socio-economic status, behavior of student, teacher aid, efficiency of principles, gender etc

RESEARCH HYPOTHESIS

- How can we predict the grades of a student in kalboard 360?
- What factors should we consider in predicting the grades?
- Which prediction model best predicts the grades of student?

DATASET DESCRIPTION

- The educational dataset used in this study is collected from learning management system (LMS) KalBoard 360.
- This system is designed to provide synchronous access to educational resources from any device with internet connection.
- The dataset is collected through 2 educational semesters and contains data of students from different origins.
- The dataset consists of 480 unique student records, 305 males and 175 females.
- Features in the dataset are classified into 3 major categories:
- A) Demographic Features
- B) Academic Background Features
- C) Behavioral Features

DATASET DESCRIPTION

Attribute	Type	Description
Gender	Nominal	Students' gender
Nationality	Nominal	Students' nationality
Place of Birth	Nominal	Students' place of birth
Educational Stages	Nominal	Educational level student belongs
Grade Levels	Nominal	Grade student belongs
Section ID	Nominal	Classroom student belongs
Topic	Nominal	Course topic
Semester	Nominal	School year semester
Parent responsible for student	Nominal	Parent responsible for student
Raised hand	Numeric	Number of times student raises his/her hand on classroom
Visited Resources	Numeric	Number of times a student visits a course content
Viewing announcements	Numeric	Number of times the student checks the new announcement
Discussion groups	Numeric	Number of times the student participates on discussion groups
Parent Answering Survey	Nominal	Parent answered the survey provided from school or not
Parent school satisfaction	Nominal	The degree of parent satisfaction from school
Student absence days	Nominal	Number of absence days for each student

METHODOLOGY

- Step 1: Data Collection and Data Preparation
- Step 2: Feature Selection
- Step 3: Clustering (k-means) and Assignment of Grades
- Step 4: Model Building
- Step 5: Parameter Tuning or Hyperparameter Optimization.

CLUSTERING AND ASSIGNMENT OF GRADES

- Clustering is done based on four features Raised Hands, Visited Resources, Announcement Views, Discussion Groups
- The clustering algorithm used is k-means (Repeated Several times).
- Entire dataset is divided into 3 clusters.
- The cluster with highest values of all four features is assigned with grade A followed by the second highest with grade B and the least with grade C.

MODEL BUILDING

- Prediction Methods Used:
- A) Support Vector Machine
- B) Decision Tree
- C) Naive Bayes
- In order to minimize the bias, the prediction methods are trained and tested using the 10-fold cross validation repeated 3 times

PARAMETER TUNING OR HYPERPARAMETER OPTIMIZATION

- Parameter Tuning is an important step for improving the algorithm performance right before presenting the results or preparing the system for production.
- It is sometimes called Hyperparameter Optimization
- Parameter Tuning can be phrased as a search problem, different search strategies can be employed to find a good and robust parameter.
- In this study, we employed Grid Search to find the hyperparameter.

GRID SEARCH

- Grid Search is an approach to parameter tuning that will methodically build and evaluate a model for each combination of algorithm parameters specified in a grid.
- It is sometimes referred as parameter sweep.
- In simple terms, it is searching through a specified subset of hyperparameter space of a learning algorithm.
- In this search, the parameter grid is fed to the respective predictive method and the method selects the best parameter with high accuracy value.

RESULTS

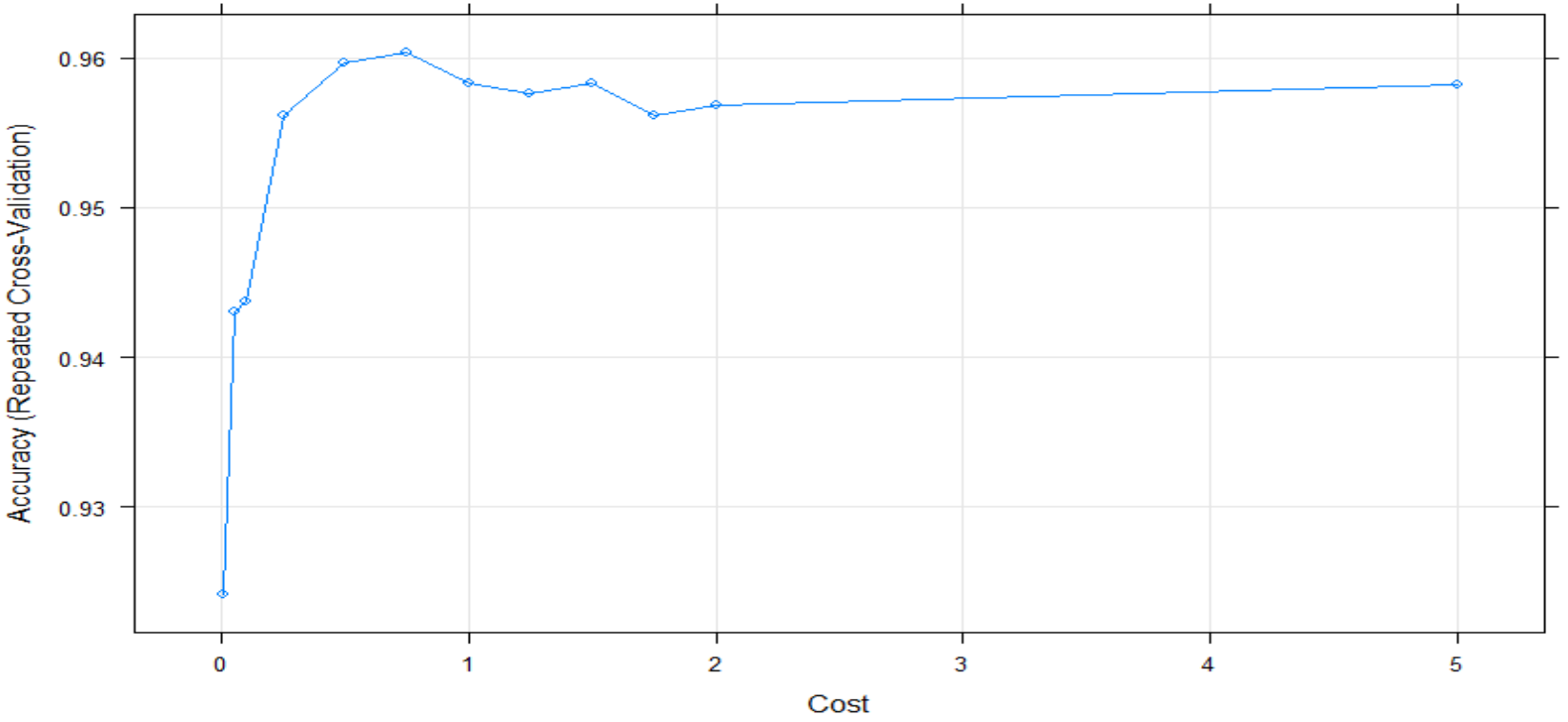
- Prediction results before parameter tuning

Prediction Method	Accuracy	Kappa statistic
SVM Linear	0.9541612	0.9305368
Naïve Bayes	0.7738843	0.6545808
Decision Tree	0.9099080	0.8632813

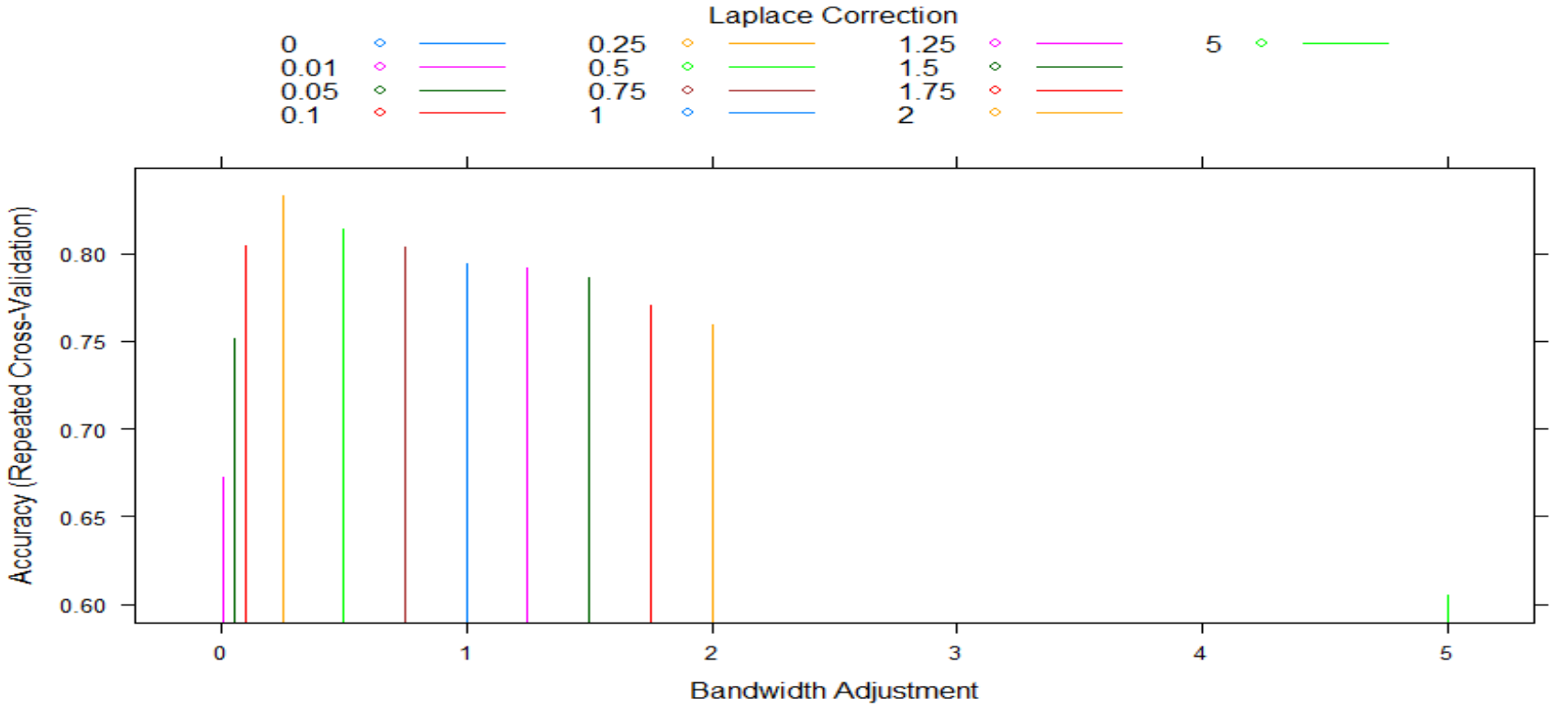
- Prediction results after parameter tuning

Prediction Method	Accuracy	Kappa statistic
SVM Linear	0.9603368	0.9398497
Naïve Bayes	0.8332935	0.7428229
Decision Tree	0.9341451	0.9004462

PARAMETER TUNING SVM



PARAMETER TUNING NAÏVE BAYES



CHALLENGES FACED

- Feature Selection
- Grid values in parameter tuning

ACKNOWLEDGEMENT

- I Sincerely thank professor Dr. Collin Lynch for his guidance and support for successful completion of this project.
- I would also like to thank my classmates for providing feedback to improve my project.

QUESTIONS?

THANK YOU