Prediction of English Language Proficiency Assessments for California (ELPAC) Scores Using Machine Learning Approach

Emma Oo
Applied Data Science Master's Program
Shiley Marcos School of Engineering / University of San Diego
eoo@sandiego.edu

Luke Awino
Applied Data Science Master's Program
Shiley Marcos School of Engineering / University of San Diego
lawino@sandiego.edu

Oscar Gil
Applied Data Science Master's Program
Shiley Marcos School of Engineering / University of San Diego
ogil@sandiego.edu

ABSTRACT

Each year, English Learners in California K-12 schools take the English Language Proficiency Assessments for California (ELPAC) Summative Assessment. Students who score an overall Level 4 on the ELPAC are up for reclassification to Fluent English Proficient. Upon reclassification, a student never has to retake the ELPAC assessment.

School districts evaluate ELPAC results upon release during the summer break to determine reclassification rates. Additionally, students with multiple years of ELPAC testing who are not up for reclassification are investigated closely to resolve their long-term English Learner status.

Our project uses five years of ELPAC data combined with student's attendance, ethnicity, age at the date of the test, day of the week when the student tested, teacher's total years of education, teacher's ethnicity, and several other features with the intent of predicting a student's overall level.

We hypothesized attendance would factor into the prediction, where poor attendance correlates to a poor overall level, while excellent attendance increases the chance of a better overall level. With attendance and other surprising features forming part of the final machine model, school districts can benefit from machine learning to help them predict ELPAC results by combining local supplemental data unavailable in the ELPAC dataset. Through predictive analytics, school districts can increase reclassification rates.

KEYWORDS

ELPAC, English Learners, CAASPP, CALPADS

1 Introduction

The English Language Proficiency Assessments for California (ELPAC) assessment helps California school districts gauge how efficient their English Learner programs are (California Department of Education, 2022). The efficiency is reflected by how many students score Level 4, resulting in reclassification to “Fluent English Proficient.”

The goal of this project is to identify features of a student profile that is most likely to succeed in the ELPAC assessment. School districts in California can benefit from having this predictive knowledge so, through formative assessments, they can identify students expected to score well on the ELPAC summative assessment.

In addition, assessments can help identify deficiencies for students who might be on the cusp of being expected to perform well. This data science project is vital because at an individual student level, it will promote student success and at a school district level, it will help gauge the effectiveness of all English learner systems in place.

2 Background

The ELPAC test, given to students whose primary language is not English, is used to measure proficiency of the English language from kindergarten (K) to Grade 12 (California Department of Education, 2022). There are initial ELPAC and summative ELPAC assessments. Initial ELPAC identifies whether students are English learners or ones fluent in English. The summative ELPAC assesses the progress of English learners in listening, speaking, writing, and reading in English. English learner students are given the ELPAC test every year. The test scores are classified into Level 1, 2, 3, and 4. Once the students score Level 4, they are reclassified as fluent English proficient (California Department of Education, 2022). The levels are broken down as:

Level 4: Well developed
Level 3: Moderately developed
Level 2: Somewhat developed
Level 1: Minimally developed

2.1 Problem Identification and Motivation

Education is vital for all communities since it can open up opportunities for all families with different backgrounds. Through education, a student from a low-income family may be able to increase their earning capability and be able to financially benefit their family, provide funds and expertise in developing their community, or be a national leader in a specialized field. Efficiency in the English language is critical to succeed in education in the United States. Most immigrant students experience the language barrier. They have to learn English in addition to other subjects as compared to native language students. Thus, the United States government realizes that not all children may have the same access and opportunities to education.

In 1965, the Elementary and Secondary Education Act (ESEA) was set into place with the goal of providing full educational opportunity to all students (U.S. Department of Education, 2022). The Elementary and Secondary Education Act (ESEA) of 1965 Title I calls for improving academic achievement of the disadvantaged, and Title III goes further into language instruction for English learners and immigrant students (U.S. Department of Education, 2022)

A key finding by the United States Department of Education (2018) recognized that, for English learner students who dropout of school, there are “limited options for improving their situation later in life” (p.12). Being able to identify which characteristics affect a student’s chances for a student to be reclassified as fluent English proficient and recognize the factors that play a role in their ability to increase their proficiency will enable students, families, and teachers to be able to collaborate on a plan for the student’s success.

2.2 Definition of Objectives

Using available data, we built a model to predict student scores on the English Learner assessment to gain an understanding of the factors impacting student success on the English Learner assessment. We used data gathered from a California school district.

The data has been deidentified to remove any potentially identifying information such as district name, schools, staff, and student names and ids. We cleaned the raw data and consolidated into a single dataset by combining datasets spanning several years with different columns. In addition to selecting the features we need; we also created the additional features by feature engineering which was mentioned in detail in the Methodology section.

With the hypothesis that students with poor attendance are likely to perform poorly on the ELPAC exam, the machine learning model predicts the student's ELPAC levels. This study can benefit from classifying the students who will not score Level 4 instead of waiting another year to take the ELPAC. For students who have been predicted as not scoring Level 4, a plan of action such as after school programs, tutoring sections, and/or guiding students on areas to focus extensively can be made to help students pass the ELPAC.

3 Literature Review (related works)

Accurately predicting student's future performance based on their ongoing academic records is crucial for effectively carrying out necessary pedagogical interventions to ensure students' on-time and satisfactory graduation (Xu et al., 2017). In the study of predicting the student performance in completing the degrees based on their current and past academic performance, an anonymized student dataset from University of California, Los Angeles (UCLA) containing pre-college traits with high school GPA and SAT scores was trained to predict the students' degree completion. The models used were Linear Regression, Logistic Regression, Random Forest, kNN, and ensemble-based progressive prediction which incorporate students' evolving performance into prediction. Among them, ensemble-based progressive prediction outperformed the best followed by Random Forest, while kNN performed the least per mean square error metrics. In addition to the outperforming model, the study also revealed the correlated predictor, SAT scores.

Similar research was conducted where the first-year student retention rates from University of Nevada, Las Vegas was predicted using the Logistic Regression, Decision Tree, Random Forest, and Support Vector Machines classifiers (Rajuladevi, 2018). The logistic regression classifier outperformed with AUC scores of 0.883 for the test dataset.

Another interesting study is the prediction of Portuguese high school student grades using the hybrid approach with classical statistical analysis and artificial intelligence (Costa-Mendes et al., 2021). A multilinear regression model was used in parallel with random forest, support vector machine, artificial neural network and extreme gradient boosting machine stacking ensemble implementations. The machine learning algorithms attain a higher level of predictive ability (Costa-Mendes et al., 2021).

Naicker et al. (2020) also used the Linear Support Vector Machines in prediction of students' performance. The linear SVM classifier was benchmarked with ten other algorithms such as coarse decision tree, medium decision tree, fine decision tree, logistic regression, Gaussian Naive Bayes, Kernel Naive Bayes, quadratic SVM, cubic SVM, fine Gaussian SVM, and medium Gaussian SVM (Naicker et al., 2020). Linear support vector machines showed superior performance in predicting student performance.

In the prediction of student performance based on personality traits and Intelligence Quotient using kNN, Naives Bayes, Random Forest, Decision Tree, and Support Vector Machine (SVM), the decision tree technique gave the best performance among the other techniques with accuracy = 0.90%, precision = 0.89%, recall = 0.90% and F1 measure = 0.89% (Samar El-Keiey et al., 2022).

4 Methodology

In this section, we discuss the steps involved from data acquisition of the raw data, exploratory data analysis, data quality, and feature engineering which led to the final output data to be used for the machine learning models.

4.1 Data Acquisition and Aggregation

Five years of ELPAC Assessment data were obtained for an elementary school district in California from the California Assessment of Student Performance and Progress (CAASPP) system. Attendance information for five school years (2017–2018 through 2021–2022) came from Student Attendance Summary (STAS) files from the California Longitudinal Pupil Achievement Data System (CALPADS). Student demographic and teacher information was obtained from CALPADS reports 8.1 and 4.4, respectively. School, student, and teacher identifiable information was de-identified. All datasets were merged into a final raw file.

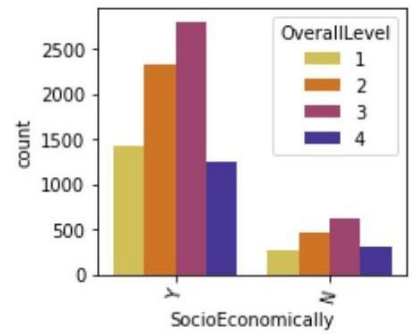
4.1.1 Exploratory Data Analysis

The raw data consists of 11,628 records with 24 attributes including information regarding academic year, de-identified school and student IDs, student's date of birth, student's age at the test date, gender, ethnicity, special education status, homeless status, socioeconomic status, test related information such as test day name, test date, overall scores, overall level, expected attendance in days and percentages, attendance in days and percentages, enrolled percentage, grade attended and grade enrolled percentages, and information related to teachers such as their gender, years of service, and ethnicity.

Figure 1 shows the students' socioeconomic status. Most of the students are from low-income families. Socioeconomic status does not appear to affect students' ELPAC scores as students from low-income families have similar ELPAC score distributions as students from other income levels.

Figure 1

Student Socioeconomic Status



4.1.2 Student Housing breakdown

Less than 10% of the student population is homeless. Table 1 shows the detailed association between housing status and ELPAC scores in percentages.

Table 1

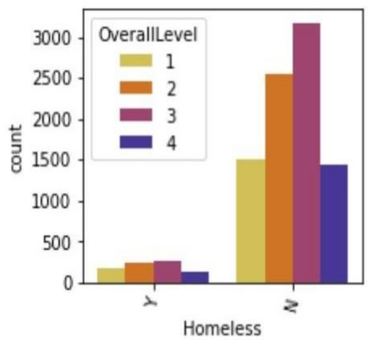
Student Housing Status

Overall Level	Homeless (%)	Not Homeless (%)
1	.106	.893
2	.086	.914
3	.926	.074
4	.923	.077

Student housing status does not appear to influence student scores as they have a similar score distribution with most of the students in both populations having an overall score of Level 3 (see Figure 2).

Figure 2

Student Housing Status

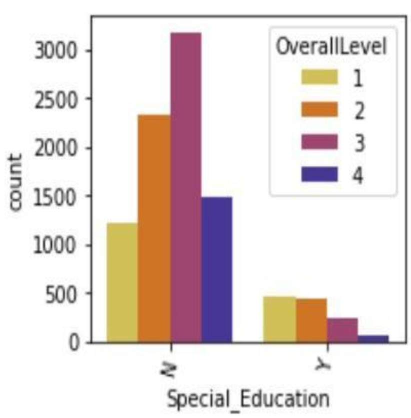


4.1.3 Special Education Status

The majority of students in special education score between Levels 1 and 2 while those without special education score Level 2 and 3. Thus, students with special education are more challenged to score at a higher level.

Figure 3

Special Education Status

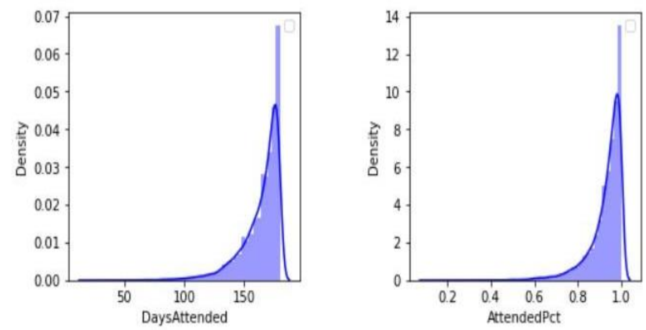


4.1.4 Attendance Distribution

Figure 4 shows the distribution of the students' attendance for the number of days attended and the percentage of days attended. The left-skewed histogram indicates most students attend school on a frequent basis indicating a high attendance rate.

Figure 4

Attendance Distribution

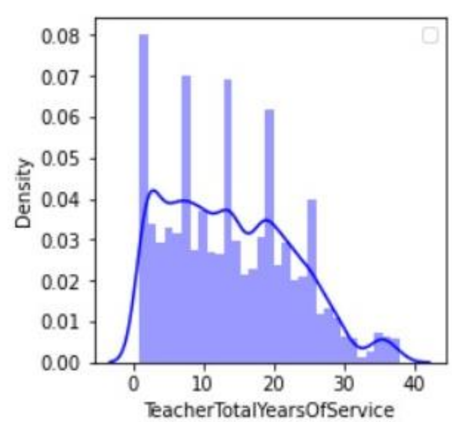


4.1.5 Teacher Years of Service

Figure 5 shows a right-skewed distribution indicating that most of the teachers are newer teachers. Most of the teachers fall between 0-20 years of service; there is a sharp decline after 20 years of service.

Figure 5

Teacher's Total Years Of Service

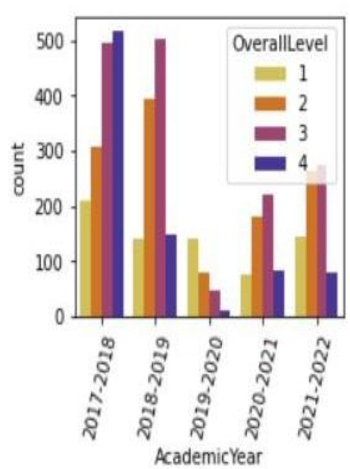


4.1.6 Student Scores by Year

During the 2017–2018 school year there were more students who achieved a Level 4 score than any other year. Since then, there was a dramatic drop off in Level 4 scores; the 2019–2020 and 2020–2021 years had the lowest number of overall students taking the test due to the COVID-19 pandemic. Among them, the 2019–2020 year had the highest proportion of students scoring Level 1 score and the lowest proportion of Level 4 scores. Thus, 2019-2020 was the year where the poor performance with lowest classification rates occurred.

Figure 6

OverallLevel By Academic Year

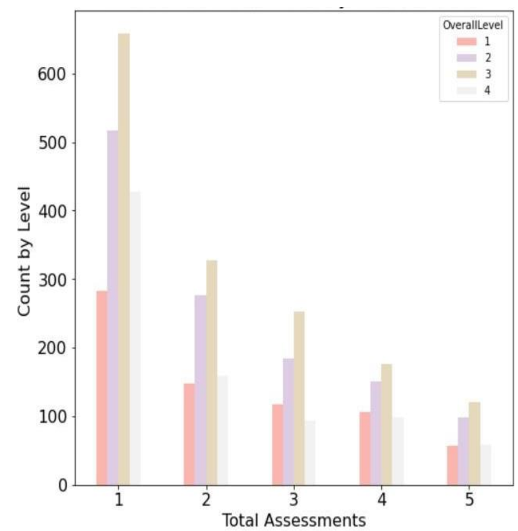


4.1.7 Total assessments by Level

Figure 7 shows the total student assessments by overall level. Across all assessment tries, most students score an overall of Level 3 most often. More students score Level 4 than Level 1 on their first and second tries. As the number of assessments increases, the number of students decreases, showing more students pass the exam on their first or second try than on their fourth or fifth try.

Figure 7

Student's Assessments By OverallLevel



4.2 Data Quality

The dataset contained some missing values. There were 2,159 missing values in the columns of "TestDayName," "TestAge," "TestDate," three missing values in "AttendedPct" and "GradeAttendedPct" columns respectively. The missing 2,159 records of "TestDayName," "TestAge," and "TestDate" were related to each other. There were 858 missing values in the year 2020–2021, 1,255 in 2019–2020, and 46 in the 2018–2019 academic year. Most missing data from 2019–2020 were due to the COVID-19 pandemic quarantine which began in March 2020. The missing records from 2020–2021 also resulted from the distance learning due to the pandemic. The records with missing data were omitted from the dataset as there was no logical imputation method for these records. The overall scores with null values were also omitted as these data were from students who took a special version of the ELPAC. After dropping these records, the data set consists of 9,467 records. The moderate class imbalance for the target classes of 'OverallLevel' was also observed.

4.3 Feature Engineering

The following features were engineered.

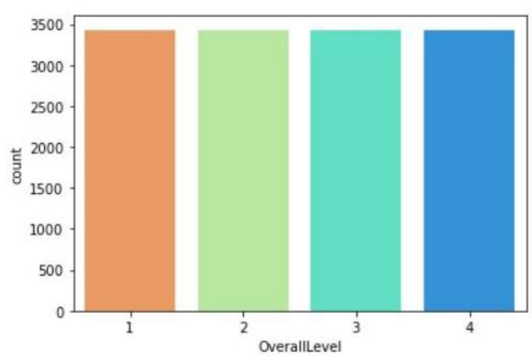
1. TestDayName: TestDayName represents the day name associated with the test date.
2. TestAge: TestAge represents the student's age at the test's date in a six-decimal precision float.
3. AttendedPct: AttendedPct reflects the percentage of each student's attendance total.
4. EnrolledPct: EnrolledPact is the percentage of total enrolled days out of 180 possible attendance days.
5. GradeEnrolledPct: GradeEnrolledPct is the student's grade level concatenated with the enrolled percentage.
6. GradeAttendedPct: GradeAttendedPct is the student's grade level concatenated with the attended percentage.
7. OverallScoreStd: OverallScoreStd is the range of overall scores by grade level that is not uniform. OverallScoreStd contains standardized overall grade-level scores so that minimum and maximum results are from 0.0 to 1.0 across all grade levels.
8. TotalAssessments: TotalAssessments contains the number of total assessments by a student, represented in the dataset.
9. TestInstance: TestInstance is the numeric representation of each ELPAC instance by student, in ascending fashion by test date.
10. Growth: Growth represents the growth in OverallLevel, comparing the OverallLevel from the most recent TestInstance with the next in the series, by student.
11. Label-encoding: Laben-encoding was performed for the categorical variables, StudentGender, StudentEthnicity, Special_Education, Homeless, SocioEconomically,TestDayName, TeacherGender, and TeacherEthnicity.

4.4 Modeling

After exploring and analyzing our dataset, we built different models to see which ones would get the best results. Before modeling the class imbalance needed to be addressed; if not addressed the models would spend more time learning the majority classes than the minority classes. Learning all the classes was important for this project especially given that passing Level 4 was in the minority class. To address the class imbalance in the target variable, the minority classes were up-sampled to balance the classes; Figure 8 shows the class distribution after balancing.

Figure 8

Distribution of Balanced Target Classes



4.4.1 Selection of modeling techniques.

Several classification models were chosen for the modeling portion, given that different models have different strengths and weaknesses, we built different models. Some models were further hypertuned to get the best results.

4.4.2 Test design (i.e., training and validation datasets)

For our dataset we had a train test split of 70% train set and 30% test. This was an appropriate split to give enough data for training and testing with 6,622 and 2,838 records, respectively. The second set of train and test data sets with standardized features were also generated to examine the performance metrics' improvements for kNN and Logistic Regression models.

4.4.3 Logistic Regression Model. The logistic regression model was chosen for its ability to make multi-class classification, as well as give an output of the important features used to make the calculation. We used several Logistic regression models, The first logistic regression model used as a baseline was on the non-standardized dataset with an accuracy of 58%. The second logistic regression model was applied to the standardized dataset with an accuracy of 85%. The third logistic regression model was a penalized logistic regression model on the standardized dataset with an accuracy of 85%, the penalty used was the L2 (Ridge Regression); the penalty prevents overfitting in the model by penalizing models with multiple features that may have collinearity (Kuhn & Johnson, 2019). The Final logistic regression model was hyper-tuned using several parameters (solvers, penalty, regularization strength), this model had an accuracy of 84%.

4.4.4 Decision Tree Model. The Decision Tree Model was chosen for its versatility in dealing with classification and regression models as well as its ability to visualize the features selected. The parameters used were max_depth of 5 and the Gini criterion was used for its ability to determine how well the splits worked the accuracy was 83%.

4.4.5 Gradient Boost Classifier Model (GBC). We used two Gradient boosted models; boosted trees work by starting off as weak learners with few features and gradually increasing features over different iterations while increasing the accuracy and reducing the errors over each iteration. The first Gradient boosted model used all the features in the dataset and had an accuracy of 96.72%. The second Gradient boosted model used the top six important features from the first Gradient boosted model and had a slight increase in accuracy 96.93%

4.4.6 k-Nearest Neighbors Model. The kNN model was chosen for its classification ability to segment similar features that are close to one another into like classes. We used two different kNN models tuned using distance parameters Euclidean and Manhattan to see which one had the better performance. The kNN model with Euclidean distance had an accuracy of 55% and the kNN model with Manhattan distance had an accuracy of 60%. using the parameters from the Manhattan model optimal parameters a kNN model was trained with an accuracy of 60%.

4.4.7 Histogram Gradient Boost Model (HGB). The Histogram based Gradient boost classifier was chosen for its similarity to the Gradient-boosted model as well as its ability for faster calculations than the Gradient boosted model, since boosted models start with fewer features and add features over the iterations, they take a long time to calculate. The Histogram based Gradient boosted model can overcome this by binning the continuous input variables resulting in the faster decision tree (Brownlee, 2021). When dealing with large datasets or if compute costs are a factor this would be an alternative to the Gradient boost classifier. This model had an accuracy of 97.46%. A second Histogram based Gradient boosted model was created using the top features, it had an accuracy of 92.92%, this model was used for creating the web application for simplicity and convenience.

4.4.8 Random Forest Classifier Random Forest was one of the models trained due to its easier usage, accuracy improvement, and insensitivity to outliers. The optimal max_depth of 13 was obtained by parallel running the accuracy of the train and test dataset from the range of 1 to 50. The Random Forest classifier was then trained with max_depth 13.

5 Results and Findings

Accuracy and F1 scores were performance metrics of choice to measure success of the trained models. The "F1 score conveys the balance between the precision and the recall" (Brownlee, 2019). Accuracy provides an average of F1 Scores, which is essential as we have four F1 scores per model for each of the four Overall levels.

The accuracy scores for all the trained models were as shown in Table 2 and Figure 9.

Table 2

Accuracy Scores of Each Model

Model	Accuracy (%)
LG	58.1
LG SCALED	84.57
PENALIZED LG SCALED	84.57
TUNED LG	84.43
KNN	60.11
DECISION TREE	82.7
RANDOM FOREST	94.12
GBC	96.72
GBC FEAT	96.93
HGB FULL	97.46
HGB PARTIAL	92.92

Figure 9

Accuracy Scores of Each Model

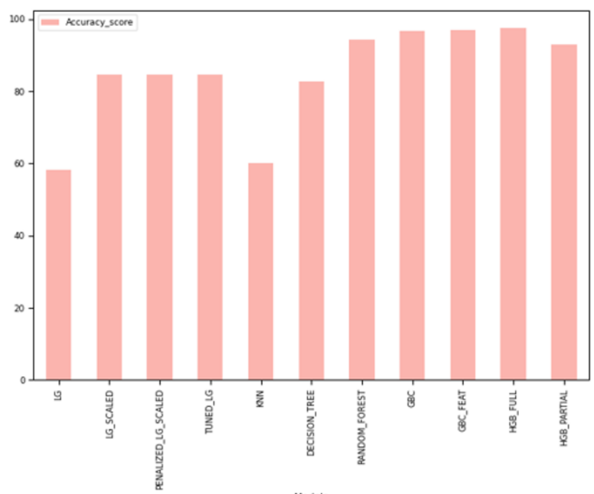


Table 3 and figure 10 below represents the f1-scores of all the trained models.

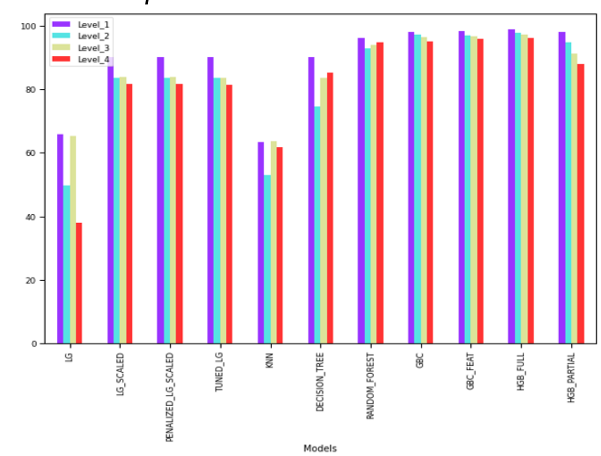
Table 3

F1 Scores of Each Model

Model	Level 1	Level 2	Level 3	Level 4
LG	65.72	49.73	65.23	37.92
LG SCALED	90.16	83.64	83.84	81.62
PENALIZED LG SCALED	90.16	83.64	83.84	81.62
TUNED LG	90.11	83.51	83.69	81.35
KNN	63.44	53.1	63.72	61.72
DECISION TREE	90.05	74.48	83.49	85.23
RANDOM FOREST	96.06	92.85	93.87	94.84
GBC	98.14	97.08	96.42	95.15
GBC FEAT	98.25	97.06	96.68	95.8
HGB FULL	98.84	97.84	97.09	96.07
HGB PARTIAL	98.05	94.63	91.15	87.89

Figure 10

F1 Score of each model



5.1 Evaluation of Results

Logistic Regression model with non-standardized data yields the lowest accuracy of 58.1% followed by kNN (60.11%), Decision Tree (82.7%), Logistic Regression with standardized dataset (84.57%), Penalized Logistic Regression with standardized dataset (84.57%), hyperparameter tuned Logistic Regression (84.43%), HGB with partial features (92.92%), Random Forest (94.82%), GBC (96.72%), GBC with top 6 important features (96.93%), and HGB with all features (97.46%). Thus, standardizing the dataset is necessary to improve the accuracy for the logistic regression model.

The lowest f1 scores were observed in the Logistic Regression model with non-standardized datasets. Secondly, kNN yields the lower f1 scores after Logistic Regression as shown in Table 5.2. The remaining models yield more than 70% of f1 scores for all the target classes. Among them, HGB with all variables yield the highest f1 scores with 98.84%, 97.84%, 97.09%, and 96.07% for Level 1, Level 2, Level 3, and Level 4, respectively.

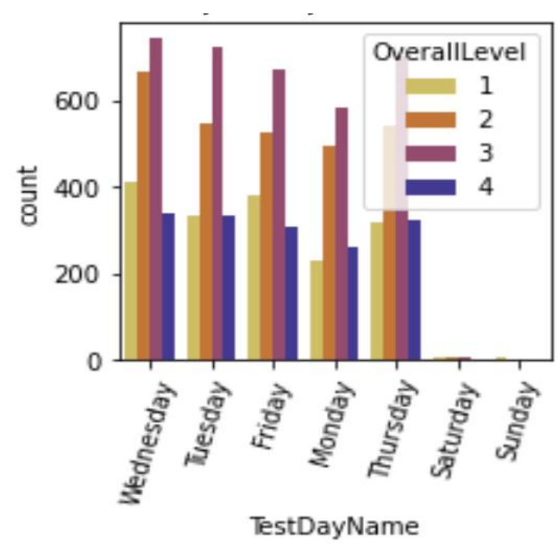
Thus, the optimal model is HGB with all features according to the accuracy and f1 scores. The HGB with partial features which contained only the school ID and student's information primarily such as student's grade level, gender, ethnicity, special education, homeless, socioeconomically, test day, overall score, expected attendance days, and days attended was also trained for the purpose of creating the simple end user web application. The accuracy remained optimal (92.93%). Thus, the web application was built using Streamlit. The end user web application can be accessed at the following:
https://oscarg-datasci-ads-599b-streamlitelpac-app-obqftk.streamlit.app/

6 Discussion

The original hypothesis considered that students with poor attendance are likely not to do well on the ELPAC Summative Assessment, scoring a low OverallLevel. The original hypothesis stands true regarding the importance of attendance, but our final model introduces other supporting features which were not on our radar initially. For example, the school of attendance weighs in on the OverallLevel, which is surprising because, typically, all schools within a district have standardized teaching practices across all schools, so one might expect the school of attendance not to make an impact on results. In addition, the day of the week the student takes the ELPAC Assessment also factors into the predicted OverallLevel, significantly impacting students whose OverallLevel results in 2 or 3, as reflected in Figure 11.

Figure 11

OverallScore By TestDayName



Our final model's results include identifiers from the features for low-income identification, whether a student is homeless, participates in special education, and socioeconomic status, as mentioned in our problem statement. Most students in the district in our model are low-income; however, the school of attendance is where actionable insight opportunities lie. In addition, a school ID played an important role in prediction of ELPAC levels. Therefore, the school district can compare ELPAC results across grade levels and schools to determine where best practices lie to then apply those practices throughout the school district.

6.1 Conclusion

In this paper, we proposed machine learning techniques for predicting the ELPAC levels for grades 0 to 6, where grade level 0 equals kindergarten. Some feature engineering such as total assessment, growth, and test instances were generated from the raw data set to monitor the student's performance. Among all the eleven models trained, gradient boosted models outperform with more than 90% of accuracy score. In addition, the web application was built using the optimal model, HGB. This web application can be used to evaluate the students' performance and where they stand with the ELPAC exam instead of waiting another academic year to find out the scores.

6.2 Recommend Next Steps/Future Studies

Future works involving the pedagogical measures can be carried out to improve the school's ELPAC passing scores. Since the school district is one of the important features in the machine learning model, schools with poor ELPAC passing rates can replicate the academic curriculum programs and attendance policy. An after-school program for the students with challenges such as homelessness, special education, and socioeconomics is suggested.

Future studies can also include exploring the relationship of ethnicities between teachers and students to determine if there is a correlation between student's success and having a teacher with a similar background.

Our recommendation for those wanting to take our model further depends on whether the model will include data for one school district or a collection of school districts. If there's a collection of school districts, we recommend creating a feature for the school district such as District_deID. Additionally, adding a feature to represent the percentage of English Learners in each school district would help the model in terms of bias as not all school districts have the same percentage of English Learners. Finally, additional analysis of results based on the growth feature may reveal that certain teachers succeed more than others in producing students who score Overall Level 4. Therefore, school districts can benefit from implementing commonalities in the practices of these particular teachers.

ACKNOWLEDGMENTS

We want to acknowledge the following organizations and websites for providing us the resources for the success of our project:

- California Assessment of Student Performance and Progress (CAASPP)
- California Longitudinal Pupil Achievement Data System (CALPADS)
- Streamlit

References:

Brownlee, J. (2019, June 20). *Classification accuracy is not enough: More performance measures you can use.* Machine Learning Mastery. https://machinelearningmastery.com/classification-accuracy-is-not-enough-more-performance-measures-you-can-use/

Brownlee, J. (2021, April 26). *Histogram-based gradient boosting ensembles in Python.* Machine Learning Mastery. https://machinelearningmastery.com/histogram-based-gradient-boosting-ensembles/

Costa-Mendes, R., Oliveira, T., Castelli, M., & Cruz-Jesus, F. (2021). A machine learning approximation of the 2015 Portuguese high school student grades: A hybrid approach. *Education and Information Technologies*, *26*(2), 1527–1547. https://doi.org/10.1007/s10639-020-10316-y

California Department of Education. (2022, November 30). *English language proficiency assessments for California (ELPAC).* https://www.cde.ca.gov/ta/tg/ep/

U.S. Department of Education. (n.d.). *Every student succeeds act (ESSA).* https://www.ed.gov/essa?src=rn

Kuhn, M., & Johnson, K. (2019). *Applied predictive modeling.* Springer.

Naicker, N., Adeliyi, T., & Wing, J. (2020). Linear support vector machines for prediction of student performance in school-based education. *Mathematical Problems in Engineering, 2020,* Article 4761468. https://doi.org/10.1155/2020/4761468

Rajuladevi, Aditya, "A Machine Learning Approach to Predict First-Year Student Retention Rates at University Of Nevada, Las Vegas" (2018). UNLV Theses, Dissertations, Professional Papers, and Capstones. 3315. http://dx.doi.org/10.34917/13568702

El-Keiey, S., ElMenshawy, D., & Hassanein, E. (2022). Student's performance prediction based on personality traits and intelligence quotient using machine learning. *International Journal of Advanced Computer Science and Applications (IJACSA), 13*(9). http://dx.doi.org/10.14569/IJACSA.2022.0130934

U.S. Department of Education. (2018, January 24). *English learner tool kit* (OELA). Retrieved November 5, 2022, from https://www2.ed.gov/about/offices/list/oela/educationandemployment.pdf

U.S. Department of Education. (2019, January 2). *Non-regulatory guidance: English learners and Title III of the Elementary and Secondary Education Act (ESEA), as amended by the Every Student Succeeds Act (ESSA).* https://www2.ed.gov/policy/elsec/leg/essa/essatitleiiiguidenglishlearners10219.pdf

Xu, J., Moon, K. H., & Van Der Schaar, M. (2017). A machine learning approach for tracking and predicting student performance in degree programs. *IEEE Journal of Selected Topics in Signal Processing*, *11*(5), 742-753. https://doi.org/10.1109/JSTSP.2017.2692560