Final Report for School District Project

Members: Abhay Kumar Singh, Mukund Srinath Heragu, Harshit Narang, Rizu Jain, Shubham Sanghavi, Sukanya Sravasti

Two-Paragraph Summary:

The stakeholder for this project is Educational leadership Research Center (ELRC), Texas A&M. They wanted to develop an AI program to determine if a school is at risk of failure in the state testing and provide this as a service to different schools. Currently, it is manually determined whether a school is at risk of failure using data from individual student scores and different variables. This process is very time-consuming. Hence, the user would benefit from an AI application, where he/she will be able to upload the individual student data for prediction and get an automated prediction from a trained model that outputs the risk of failure for specific schools.

For this semester, the customer wanted us to focus on a "proof of concept" to implement a machine learning (ML) model to predict student test scores for three different grades (3,4 and 5) and two subjects (Math and Reading). We were provided with raw data from Carmichael school that had individual-level student data on past test scores, and other features for prediction. After cleaning and preprocessing the data, we used the following available features for training our ML model to predict student scores: previous year test score, past two years mock test scores, ethnicity, socioeconomic status, enrollment duration, English proficiency. We used predictions from our ML model to produce a certificate for Carmichael school using the accountability rating system specified by the Texas Education Agency.

Description of all user stories:

- Grade 3 - Preprocess the student data mock test scores and final test scores for the machine learning models to predict the student scores (3 points): We cleaned the raw data for grade 3 Reading and Math subjects before we implemented the ML model. The format of the data was different for different grades and subjects so there was some significant amount of data cleaning. **Status:** Done
- Grade 4 - Preprocess the student data mock test scores and final test scores for the machine learning models to predict the student scores (3 points): We cleaned the raw data for grade 4 Reading and Math subjects before we implemented the ML model. The format of the data was different for different grades and subjects so there was some significant amount of data cleaning. **Status:** Done
- Grade 5 - Preprocess the student data mock test scores and final test scores for the machine learning models to predict the student scores (3 points): We cleaned the raw data for grade 5 Reading and Math subjects before we implemented the ML model. The format of the data was different for different grades and subjects so there was some significant amount of data cleaning. **Status:** Done

- Develop individual models for 5th grade math and reading (2 points): We trained a linear regression model on the individual-level data, and then used the model to predict student scores for the test dataset. An example of the prediction results that compared actual student scores to predicted student scores is shown below.

Actual STAAR percent score	Actual Approaches	Actual Meets	Actual Masters	ML Predicted Final Percent Score	ML Approches	ML Meets	ML Masters	Error
72.00%	TRUE	TRUE	FALSE	71.60%	TRUE	TRUE	FALSE	0.40%
78.00%	TRUE	TRUE	FALSE	66.50%	TRUE	FALSE	FALSE	11.50%
89.00%	TRUE	TRUE	TRUE	77.50%	TRUE	TRUE	FALSE	11.50%
36.00%	FALSE	FALSE	FALSE	51.30%	TRUE	FALSE	FALSE	15.30%
97.00%	TRUE	TRUE	TRUE	90.30%	TRUE	TRUE	TRUE	6.70%
72.00%	TRUE	TRUE	FALSE	52.50%	TRUE	FALSE	FALSE	19.50%
86.00%	TRUE	TRUE	TRUE	80.90%	TRUE	TRUE	FALSE	5.10%
75.00%	TRUE	TRUE	FALSE	73.90%	TRUE	TRUE	FALSE	1.10%
31.00%	FALSE	FALSE	FALSE	39.60%	FALSE	FALSE	FALSE	8.60%
72.00%	TRUE	TRUE	FALSE	70.80%	TRUE	TRUE	FALSE	1.20%
97.00%	TRUE	TRUE	TRUE	100.60%	TRUE	TRUE	TRUE	3.60%
75.00%	TRUE	TRUE	FALSE	71.70%	TRUE	TRUE	FALSE	3.30%
69.00%	TRUE	FALSE	FALSE	63.80%	TRUE	FALSE	FALSE	5.20%
89.00%	TRUE	TRUE	TRUE	89.50%	TRUE	TRUE	TRUE	0.50%
42.00%	FALSE	FALSE	FALSE	43.60%	FALSE	FALSE	FALSE	1.60%
64.00%	TRUE	FALSE	FALSE	52.40%	TRUE	FALSE	FALSE	11.60%
39.00%	FALSE	FALSE	FALSE	52.00%	TRUE	FALSE	FALSE	13.00%
81.00%	TRUE	TRUE	FALSE	80.50%	TRUE	TRUE	FALSE	0.50%
78.00%	TRUE	TRUE	FALSE	65.20%	TRUE	FALSE	FALSE	12.80%
75.00%	TRUE	TRUE	FALSE	73.70%	TRUE	TRUE	FALSE	1.30%
53.00%	TRUE	FALSE	FALSE	63.40%	TRUE	FALSE	FALSE	10.40%

  Status: Done
- Machine learning model tuning and training (2 points): We trained a random forest classifier using grid search for hyper-parameter tuning for the grade 5 prediction for Reading and Math. However, for the customer presentations, and rest of the grade predictions, we stuck with the Linear Regression model and used it to present our results. **Status:** Partially done since we did not implement it beyond grade 5.
- Develop individual models for 4th grade math and reading (2 points): We trained a linear regression model on the individual-level data, and then used the model to predict student scores for the test dataset. **Status**: Done
- Develop individual models for 3rd grade math and reading (2 points): We trained a linear regression model on the individual-level data, and then used the model to predict student scores for the test dataset. **Status**: Done
- Compare ML predicted scores with the customer's current method of prediction (3 points):The ML prediction results were then compared to the current method the client uses to classify different students based on their test scores. An example of the comparison is shown below. The comparison shows promising results as the ML performs better than the expert. However, there are a couple of caveats here. In reality, the expert uses many parameters than the ones that they were allowed to use and hence their overall predictions would be much better than the ML algorithm. Also, the ML used the students from the same batch. However, in reality data will be available for the previous batch. **Status**: Done

Actual	AI predictions	Expert Predictions
Masters	Masters	Masters
Did not meet	Did not meet	Did Not Meet
Masters	Masters	Masters
Approaches	Approaches	Meets
Did not meet	Did not meet	Approaches
Did not meet	Did not meet	Did Not Meet
Approaches	Approaches	Approaches
Meets	Approaches	Approaches
Approaches	Approaches	Meets
Masters	Masters	Masters
Masters	Meets	Meets
Did not meet	Approaches	Meets
Did not meet	Did not meet	Approaches
Masters	Masters	Meets
Approaches	Did not meet	Meets
Masters	Meets	Meets
Masters	Masters	Masters
Approaches	Meets	Meets
Meets	Meets	Masters
Meets	Masters	Masters
Masters	Masters	Masters
Approaches	Meets	Meets
Did not meet	Did not meet	Did Not Meet
	-4	-4
	3	11

- Convert predicted scores to domain 1 and domain 2B (3 points). We converted the predicted scores to the intermediate domain scores 1 and 2B that the Texas Education Agency uses in their calculation to generate a school certificate, and compared them to the actual domain scores in those domains. See below. **Status:** Done

School certificate Domain 1 and 2B

ML predictions v/s Actual

Domain 1		
	Scaled score	78
	Grade	C
Domain 2B		
	Scaled score	85
	Grade	B

Component Score	Scaled Score	Rating
	86	B
	76	C
48	76	
	90	A
82	90	A
48	84	B
79	78	C

- Convert predicted student scores to domain 3 (3 points). We converted the predicted scores to the intermediate domain 3 scores that the Texas Education Agency uses in their calculation to generate a school certificate, and compared it to the actual domain score in that category . See below. **Status:** Done.

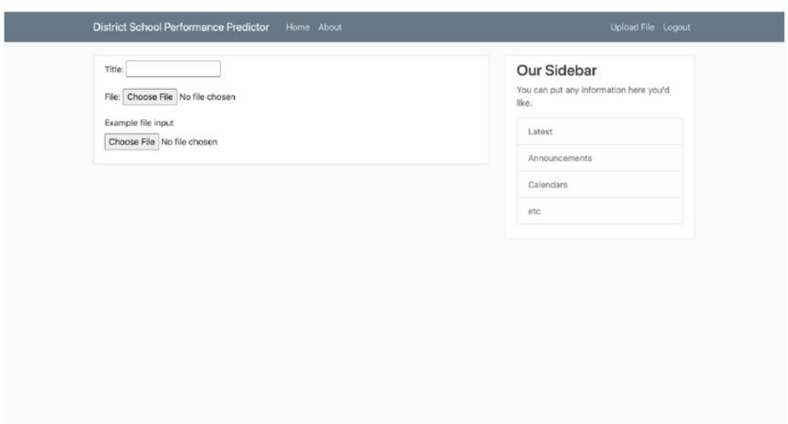
Domain 3: Actual v/s Predicted

	All Students	African American	Hispanic	White	American Indian	Asian	Pacific Islander	Two or More Races	Econ Disadv	EL (Current & Monitored)+	Special Ed (Current)
Academic Achievement Status											
ELA/Reading Target	44%	32%	37%	60%	43%	74%	45%	56%	33%	29%	19%
Target Met	N	N	Y			Y			Y	Y	N
% at Meets GL Standard or Above	41%	30%	38%	*	-	81%	-	*	40%	40%	4%
# at Meets GL Standard or Above	121	20	74	*	-	25	-	*	105	72	1
Total Tests (Adjusted)	297	66	197	*	-	31	-	*	261	179	27
Math Target	46%	31%	40%	59%	45%	82%	50%	54%	36%	40%	23%
Target Met	Y	Y	Y			Y			Y	Y	N
% at Meets GL Standard or Above	54%	39%	51%	*	-	94%	-	*	54%	55%	7%
# at Meets GL Standard or Above	159	26	101	*	-	29	-	*	141	99	2
Total Tests (Adjusted)	297	66	197	*	-	31	-	*	261	179	27

	All Students	African American	Hispanic	White	American Indian	Asian	Pacific Islander	Two or More Races	Econ Disadv	(Current & Monitored)+	Special Ed (Current)
Academic Achievement Status											
ELA/Reading Target	44%	32%	37%	60%	43%	74%	45%	56%	33%	29%	19%
Target Met	Y	N	Y	Y		Y			Y	Y	N
% at Meets GL Standard or Above	47.37%	30.77%	43.24%	100%	-	100%	-	*	44%	40.74%	0%
# at Meets GL Standard or Above	27	13	16	1	-	6	-	*	22	11	0
Total Tests (Adjusted)	57		37	1	-	6	-	*	50	27	3
Math Target	46%	31%	40%	59%	45%	82%	50%	54%	36%	40%	23%
Target Met	Y	Y	Y			N			Y	Y	N
% at Meets GL Standard or Above	50%	41.67%	48.83%	*	-	60%	-	100%	54%	48.57%	0%
# at Meets GL Standard or Above	31	5	21	*	-	3	-	2	27	17	0
Total Tests (Adjusted)	62	12	43	*	-	5	-	2	50	35	5

* Only part of the certificate shown.

- As a school personnel I should be able to access the AI application website (2 points): We developed an initial website as an example of what the user interface will look like from the school personnel's perspective. Please see below the Lofi-UI for the website that shows what the user sees after logging and getting the option to upload a csv file.



Status: Done(basic website template created)

- As a new school personnel, I should be able to register as a new user on the AI application website (2 points): The user is able to register as a new user in this website. Status: Done

- As a school personnel, I want to login to the AI application using my registered User ID and Password (2 points): The user is able to login to the AI application using the registered User ID and password. **Status**: Done
- As a school personnel, once I have logged in to the AI application, I want to upload a csv file with all the student data required for the prediction (2 points). **Status**: Not Done. We focused more on developing the ML model for student score predictions since the customer wanted us to focus on it before implementing the website.
- As a school personnel, once I have uploaded the student data, I expect the trained model to perform the prediction (3 points). **Status:** Not done. We implemented a variation of this where the trained model does perform the prediction but there was no website.
- As a school personnel, I expect to download the detailed data of the prediction results in the form of a csv file (3 points). **Status**: Not done. We implemented a variation of this where the prediction results are outputted into a csv file but there was no website.
- As a school personnel, I should be able to log out of the AI application and get a confirmation for the same (3 points). The website allows the user to log out. **Status:** Done.
- Make the platform extensible for adding High school in the future. **Status**: Not done
- Make the ML pipeline extensible to add new data columns in future. **Status**: Not done
- Pass the uploaded CSV to data preprocessor. **Status**: Not done
- Convert predicted scores to domain 2A. **Status**: Not done

Members who held team roles:

Abhay Kumar Singh - Developer
Mukund Srinath Heragu - Developer
Harshit Narang - Scrum Master
Rizu Jain - Developer
Shubham Sanghavi - Developer
Sukanya Sravasti - Product Owner

Points completed for each iteration

- **Iteration 0:** For this iteration, we set up our first meeting with the customer to get the overall big picture, introduction to the project and create customer stories. After that, we created team roles within the team. We then sketched low-fi user interfaces and storyboards.
- **Iteration 1:** For this iteration, we completed 4 user stories (9 points). We developed a Django application and deployed it to Heroku. The application is the website that allows school personnel to register and login using their registered credentials. After logging in, the user will get an option to upload a csv file that should contain the individual-level student data that the ML model will use to predict student scores.

- **Iteration 2:** A total of 2 user stories (5 points) were implemented in this iteration. First, we preprocessed the raw individual-level student data that contained student test scores and different features for prediction for grade 5 Reading and grade 5 Math. The data was in different csv files and was in different formats for different grades, so a significant amount of time was spent on understanding the data from the information the customer had given us. After cleaning the data, one final dataset was produced that contained all the student features for the ML prediction. For the second user story, we used the final preprocessed dataset to predict student scores for grade 5 Reading and grade 5 Math. We trained the data using a Linear Regression model first. Then, we trained a random forest classifier using grid search for hyper-parameter tuning. The predicted results on the test dataset were outputted in a csv file and compared with actual student scores to assess the accuracy of the predictions.

- **Iteration 3:** A total of 3 user stories (8 points) were implemented in this iteration. First, we preprocessed the raw data for grade 4 Reading and grade 4 Math in a similar way we did in iteration 2. Secondly, a Linear Regression model was trained on the data and student scores were predicted on the test dataset. and outputted in a csv file to assess accuracy compared to actual student scores. Our third user story consisted of comparing our predicted student classifications from the ML model to the current prediction method the customer uses to classify students based on their own method of prediction. We sent the customer the data that we used to predict student scores for the subset of students for grade 4 Math and Reading. This data contained all the variables that we used in our ML model except the actual final score that we used in the training phase of our model. We then compared the results of our ML predicted student classifications with the customer predicted student classifications.

- **Iteration 4:** A total of 4 user stories (11 points) were implemented in this iteration. First, we cleaned and pre-processed the data for grade 3 Reading and Math. Secondly, we predicted scores for these students. Our third and fourth user stories consisted of converting the predicted student scores to intermediate category scores that the Texas Education Agency uses to produce a school certificate. The TEA uses three different accountability domains to evaluate student performance in a school and issue a school certificate: 1) Student Achievement 2) School Progress 3) Closing the Gaps (Detailed information can be found here: https://tea.texas.gov/sites/default/files/Chapter%201%202019%20Accountability%20Overview.pdf). We converted the predicted student scores from the ML model to each of these 3 domains using the calculation provided by the TEA.

List each customer meeting dates and what you did:

● September 23, 2020 (4:30 - 6:30 pm): We met with our customer to get an overview and initiation of the project. We got an introduction to the project and what the customer expectations were.

● October 12,2020 at (7:00 pm - 9:00 pm): We met with the customer to clarify our project deliverables for this semester, and discuss our proposed user stories and their implementation plan for this project.

● October 26, 2020 at (6:00 pm to 7.45 pm): In this meeting, we presented the website that we deployed that will allow the school personnel to sign up and login using their credentials, after which they will be presented with the option to upload a csv file. While we were showing the format of the dataset after the preprocessing, it was found that the dataset we received was corrupted. So, further discussion on data cleaning was halted until the client got back to us with new data next week.

● November 2, 2020 from 5:00 pm to 7:00 pm: We received revised data from the customer that was corrupted earlier, and clarified further questions on the data. The customer informed us about their preference to focus on implementing a decent machine learning model over the GUI and web application since they needed to ensure that the ML model will produce similar student score predictions as their current prediction method. Hence, we reprioritized our goals for the semester to focus more in the ML model over the GUI. We spoke with Dr. Walker about this, and he agreed with our goals.

● November 9, 2020 from 5:00 pm to 7:00 pm: We presented results from our ML model for grade 5 Reading and grade 5 Math using the Linear Regression model. The accuracy of both these test scores was around 80%. The customer wanted us to compare their current method of human prediction with what the ML model was predicting so we decided to aim for that for the next meeting for grade 4 Reading and Math.

● November 16, 2020 from 5:00 pm to 7:00 pm: We sent the customer the data that we used to predict student scores for the subset of students for grade 4 Math and Reading. This data contained all the variables that we used in our ML model except the actual final score that we used in the training phase of our model. We then compared the results of our ML predicted scores with the customer predicted scores and actual final scores. For both Reading and Math, the ML model classified students in their proper categories with less error compared to the client's current prediction method. For example, for grade 4 Math, the ML model classified 65% students in their correct categories, whereas the client's current method of classification classified 35% of the students in their correct categories.

● November 23, 2020 from 5:00 pm to 7:00 pm: We converted the predicted student scores to two domains (1 and 2B) that the Texas Education Agency uses to assess school performance. Our results for these two domains based on the predicted student scores were close to the actual domain scores for the school.

● December 2, 2020 from 5:00 pm to 7:00 pm: We showed the client results for the remaining domain (3) that the TEA uses to assess school performance. We clarified the final format of the project code that the client would like the submission.

Explain your BDD/TDD process: We followed a BDD protocol where we incorporated weekly feedback from the client into our project goals. The client emphasized that they wanted a "proof of concept" first that an ML model could replace their current prediction method. This led us to focus more on implementing an ML model for predicting student scores instead of focusing on the GUI and web application. The client was satisfied with the proof of concept and its results.

Links to your Pivotal Tracker and GitHub repo:

GitHub
https://github.com/harshit08173/CSCE-606-TeamMarsshMellow

Pivotal Tracker
https://www.pivotaltracker.com/n/projects/2469696

Links to your poster and demo videos:

Demo:

https://vimeo.com/487498656

Poster:

https://github.com/harshit08173/CSCE-606-TeamMarsshMellow/blob/main/Documentation/Fall%202020/Poster%20and%20Project%20Demo.pptx