**Comparing Standardized Test Scores**

Lauren Lee

1. Introduction and Background

Every state is required to implement a standardized test. There are two portions to the tests. One portion mainly focuses on the English Language and tests on vocabulary, reading comprehension, and sentence structures. The next section focuses on math and sciences. Each grade level progresses and gets more challenging. The purpose of this project was to see if there were major deviations between scores per state.

1. Motivation for Project

The motivation for this project was to determine whether states with different percentages of poverty have major deviations within their standardized test scores. Each state has the liberty to declare their own standardized tests. This research focused on states who required ELA and Math tests. This research can help school leaders and state officials discover disadvantages between school groups. This discovery will help pinpoint the source of any disadvantages and aid in eliminating any prejudices. Research like this is crucial to promoting equal learning for all student groups to create a more diverse future within schools and professions. There is a stark disparity between household incomes and the top one percent’s earnings. In my opinion, this can correlate with schooling and opportunities student groups are given. There is very little to no representation of African Americans, Hispanics, American Indians, Asians, and Multiracial groups in the top one percent [1]. It is time to diversify the United States wealth. To start, we need to look at success in school.

1. Related Coursework and Datasets Used

Currently, there is little to no published research on this topic. This could be because state standardized tests vary, but conclusion can still be drawn. To pick the states to draw conclusions from, I referenced the Nation’s Report Card which displays state performance and demographics of each state. From there, I decided to pick two states who had varying poverty percentages. Oregon ranking as one of the lowest percentages of poverty and Mississippi ranking as one of the highest. During development, I referenced blogs to aid in classifying the features of the dataset and explaining the outcomes. The datasets present in the research were found from Oregon’s state government site and Mississippi’s state government site. Both consisted of ELA and Math results per student group.

The Oregon dataset gave details of each school, their district, the student groups present, the grade of the students, number of students per group, number of students who scored proficiently (levels 3 and 4), and then broken each category down future into the number in level 1, number in level 2, number in level 3, and number in level 4. In the Oregon dataset there are 7 student groups present including, Asian, Black/African American, Hispanic/Latino, American Indian/Alaskan Native, Multi-Racial, Pacific Islander, and White.

The Mississippi dataset gave details of each school district, the student groups present, the number of students surveyed, and the percent of the students who scored proficiently (levels 3 and 4). From there, I added a column which determined the number of students who scored proficiently based on the total number and the percent columns. In total, there are 6 student groups represented including, White, African American, Hispanic, Asian, Multiracial, and Native American/Pacific Islander.

1. Problem

The problem faced is whether there will be a major difference in numbers of proficiency per student group. First a comparison is drawn for the state and then each state is compared to determine if poverty levels could be a source for the problem as well.

1. User

Those who will benefit from this research are school leaders and state officials. With that also allows parents and guardians to see if their student is at a disadvantage. Students will ultimately profit off this research when the disadvantages are pointed out and fixed accordingly.

1. AI Method

The AI Methods used in this research were logistic regression and random forest to classify the data, and then LIME to explain the findings. The model uses 70% of the data to train the classifiers and 30% of the data to test. The logistic regression classifiers were 61% accurate on the training data and 64% accurate on the test data for Oregon ELA data. For the random forest classifiers were 63% accurate for the training data and 66% accurate for the test data. For the Oregon Math set, the logistic regression classifier was 59% accurate for the training data and 55% for the test set. Lastly, the random forest classifier was 61% accurate for the training set and 66% for the test set.

For the Mississippi dataset the logistic regression classifiers were 74% accurate on the training data and 77% for the test data. The random forest classifiers were 76% accurate for the training data and 78% for the test data. As you can tell, the classifiers for Mississippi were much more accurate than those for Oregon.

The LIME explanation predicts how student groups will perform per district.

1. Evaluation

First, the data was cleaned for any null values. Since the datasets were robust, I was able to completely remove any lines with null values. Next, I made some simple graphs to compared total populations and their distribution over the levels. The first graph shown here is the Oregon ELA total population distributed over the 4 levels:

Chart, bar chart

Description automatically generated

As you can tell, the majority of student groups fell in level 1. This next graph shows the Oregon Math total population distribution over the 4 levels:

Chart, bar chart

Description automatically generated

Seen in this graph, the majority is still in level 1. The performance in level 4 is much lower in the math test than the ELA per student group. Now that we know how the total population is distributed, we will look further into each student group’s performance per test.

The graphs created for the Oregon ELA distribution can be found here:

Graphical user interface

Description automatically generated

This first graph looks at level 1 distribution specially. Each dot represents a different school. From this graph we can tell how many schools were represented in level 1 and more specifically how many students per student group scored in the level 1 range. As you can tell, the American Indian/Alaskan Native student group is very slightly represented. Same with the multi-racial and Asian student groups. This next graph breaks down each student group on a numeric scale, regardless of school:

Chart, bar chart

Description automatically generated

In this graph the x-axis represents each student group, American Indian/Alaskan Native (0), Asian(1), Black/African American (2), Hispanic/Latino (3), Multi-Racial (4), Total Population (5), and White (6). It is much easier to see the White and Hispanic/Latino populations were much more predominant in this category. This could be because of the lack of representation from the American Indian/Alaskan Native (0), Asian(1), Black/African American (2), and Multi-Racial (4) groups as a whole in the dataset. The next section we will look at is level 2:

A picture containing text, screenshot

Description automatically generated

This graph shows the distribution in level 2 for each school their student groups. Like the graph before, this also shows the number per student group per school. The American Indian/Alaskan Natives group is still sparse. With a little more representation from the Multi-Racial and Asians groups. This graph also shows the most representation for the African American group that we will see. White and Hispanic groups still dominating. The next graph looks more in-depth at the specific numbers:

Chart, histogram

Description automatically generated

Now there is a little more representation from the Asian, African American, and Multi-Racial groups, but not much. Again, this may have to do with the low total representation. The next graphs are the same but correlate with level 3 and 4:

Graphical user interface, application

Description automatically generated

Chart, histogram

Description automatically generated

Graphical user interface, application

Description automatically generated

Chart, histogram

Description automatically generated

In these two levels, this is the most Asian and Multi-Racial representation. The White group is still dominate. The African American student group has little representation in any level, but the least in levels 3 and 4. After looking at these graphs, we can get a better graph of distribution and begin to draw conclusions. The last graph to look at is a correlation map:

A picture containing background pattern

Description automatically generated

This map shows correlations between all the data given through the set. The most interesting portions are the correlations between each student group and their overall proficiency

Next comes the classification. First, I split up the student groups into their own columns. Next, I wanted to see overall how many groups scored proficiently (levels 3 and 4) and how many did not. Those are as follows:

Proficient: 1006

Not: 762

The data was able to assign proficiency by adding a new column and assigning values 0 and 1 based on percent proficient. If the percentage of students was equal to or more than 50%, then group was given a 1, else the group was given a 0. This allowed me to train the classifiers and predict the percent of proficiency per student group. As stated previously, I used logistic regression and random forest to classify the student groups.

The first classification method used was logistic regression. The classifiers are trained on each student group and their expected proficiency. For the Oregon ELA, the logistic regression classifiers for the test data and training data were 64% and 66% accurate. I then continued in using the LIME method as explanation. Here is an example of an explanation given:

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | District | … | Number Proficient | … | Number of Participants | Overall Proficient? | … | Student Group\_5 | … |
| 6450 | David Douglas SD 40 | … | 13 | … | 74 | 0 | … | 1 | … |

Graphical user interface

Description automatically generated

From this random data used, the explanation is saying that at this school, the only student group that overall would score proficient would be the white student group. This could because of the lack of representation from the other student groups, but that is the conclusion that LIME draws.

The second classification method used was random forest. The classifiers are trained on each student group and their expected proficiency. For the Oregon ELA, the logistic regression classifiers for the test data and training data were both 64% accurate. I then continued in using the LIME method as explanation. Here is an example of an explanation given:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | District | … | Number Proficient | … | Number of Participants | Overall Proficient? | … | Student Group\_6 |
| 5445 | Corvallis SD 509J | … | 27 | … | 45 | 1 | … | 1 |

Chart, waterfall chart

Description automatically generated

This data is a much more evenly distributed than the previous. Student groups are split half and half on proficiency. This explanation is encouraging that the classifiers are 100% completely bias towards the White student group. There is still bias though. We can see that based on the classifiers, LIME predicts that within this school groups 2, 3, and 6 will all perform proficiently.

Next, we move onto the Oregon Math portion of the research. The graphs created for the Oregon Math distribution can be found here:

Graphical user interface, application

Description automatically generated

Chart, histogram

Description automatically generated

As touched on in the Oregon ELA section, the first graph relates the school, student group, and number of states and the second maps out the total number of students who ranked in level 4 per group. The white group is still dominate but Asian and Hispanic groups make more of an appearance than previous levels. This was a similar trend in the ELA level 4 graph as well.

Graphical user interface, application

Description automatically generated

Chart, histogram

Description automatically generated

These graphs look at the distribution among level 3.

Application

Description automatically generated with medium confidence

Chart, histogram

Description automatically generated

These graphs look at the distribution among level 2.

Graphical user interface, application

Description automatically generated

Chart

Description automatically generated

These graphs look at the distribution among level 1.

Here is the correlation graph for the Oregon Math dataset:

A picture containing background pattern

Description automatically generated

Next, comes the classification for Oregon Math. First, I split up the student groups into their own columns. Next, I wanted to see overall how many groups scored proficiently (levels 3 and 4) and how many did not. Those are as follows:

Proficient: 1013

Not: 736

This set of data shows great performance with more than 50% scoring in levels 3 and 4.

The first classification method used was logistic regression. The classifiers are trained on each student group and their expected proficiency. For the Oregon ELA, the logistic regression classifiers for the test data and training data were 62% and 63% accurate. I then continued in using the LIME method as explanation. Here is an example of an explanation given:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | District | … | Number Proficient | … | Number of Participants | Overall Proficient? | … | Student Group\_6 |
| 9228 | Gaston SD 511J | … | 4 | … | 15 | 0 | … | 1 |

Chart, waterfall chart

Description automatically generated

All the explanations that have been given have been very different. Which is great to see. We can see that based on the classifiers, LIME predicts that within this school groups 1, 2, 3, 4, and 6 will all perform proficiently.

The next classification method used was random forest. The classifiers are trained on each student group and their expected proficiency. For the Oregon ELA, the logistic regression classifiers for the test data and training data were 62% and 63% accurate. I then continued in using the LIME method as explanation. Here is an example of an explanation given:

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | District | … | Number Proficient | … | Number of Participants | Overall Proficient? | … | Student Group\_5 | … |
| 4658 | Central SD 13J | … | 30 | … | 73 | 0 | … | 1 | … |

Chart, waterfall chart

Description automatically generated

This example predicts that from this class, student groups 1, 4, and 6 will score in the proficient range.

Next, we’ll analyze the Mississippi data. The graphs created for the Mississippi ELA distribution can be found here:

Chart, scatter chart

Description automatically generated

Here is the correlation graph for the Mississippi ELA dataset:

A picture containing table

Description automatically generated

Next, comes the classification for Mississippi ELA. First, I split up the student groups into their own columns. Next, I wanted to see overall how many groups scored proficiently (levels 3 and 4) and how many did not. Those are as follows:

Proficient: 189

Not: 562

It’s interesting that 25% of the total population scored in the proficient range. Compared to the Oregon ELA results that’s a major difference in results. Could the conclusion be drawn that higher percentages of poverty result in lower performance? This question will need to be researched further with more states.

The first classification method used was logistic regression. The classifiers are trained on each student group and their expected proficiency. For the Mississppi ELA, the logistic regression classifiers for the test data and training data were 77% and 76% accurate. I then continued in using the LIME method as explanation. Here is an example of an explanation given:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | District | … | Number Proficient | … | Number of Participants | Overall Proficient? | … | Student Group\_6 |
| 9228 | Gaston SD 511J | … | 4 | … | 15 | 0 | … | 1 |

Chart, funnel chart

Description automatically generated

This explanation is interesting. It concludes that from this district, all the student groups will perform proficiently. I find that explanation very interesting especially because the given example group did not score proficiently overall. This could be an error in the explanation.

The next classification method used was random forest The classifiers are trained on each student group and their expected proficiency. For the Mississippi ELA, the logistic regression classifiers for the test data and training data were 77% and 77% accurate. I then continued in using the LIME method as explanation. Here is an example of an explanation given:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | District | 2019 Count | … | Number Proficient | Overall Proficient? | Subgroup\_1 | … |
| 262 | Greenville Public Schools | 2315 | … | 540 | 0 | 1 | … |

Chart, waterfall chart

Description automatically generated

LIME for this example concluded that within the Greenville Public School district, groups 2, 3, 4, and 6 will score proficiently. Groups 1 and 5 will not. The model’s prediction probabilities are 77% for levels 1 and 2 and 23% for levels 3 and 4.

The graphs created for the Mississippi Math distribution can be found here:

Chart, histogram

Description automatically generated

Here is the correlation graph for the Mississippi Math dataset:

Calendar

Description automatically generated with medium confidence

Next, comes the classification for Mississippi Math. First, I split up the student groups into their own columns. Next, I wanted to see overall how many groups scored proficiently (levels 3 and 4) and how many did not. Those are as follows:

Proficient: 267

Not: 482

The first classification method used was logistic regression. The classifiers are trained on each student group and their expected proficiency. For the Mississppi Math, the logistic regression classifiers for the test data and training data were 69% and 72% accurate. I then continued in using the LIME method as explanation. Here is an example of an explanation given:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | District | 2019 Count | … | Number Proficient | Overall Proficient? | … | Subgroup\_3 | … |
| 88 | Canton Public School District | 236 | … | 93 | 0 | … | 1 | … |
|  |  |  |  |  |  |  |  |  |

Chart, waterfall chart

Description automatically generated

This explanation concludes that only subgroup 6 will perform in the proficient levels.

The next classification method used was random forest The classifiers are trained on each student group and their expected proficiency. For the Mississippi Math, the logistic regression classifiers for the test data and training data were 71% and 65% accurate. I then continued in using the LIME method as explanation. Here is an example of an explanation given:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | District | 2019 Count | … | Number Proficient | Overall Proficient? | Subgroup\_1 | … |
| 148 | Coffeeville School District | 273 | … | 95 | 0 | 1 | … |

Chart, waterfall chart

Description automatically generated

This example concludes that subgroups 2, 3, 4, and 6 will perform in levels 3 and 4. Subgroups 1 and 5 will not.

1. Human-AI Issues

Some of the major trust issues that this model will face is bias towards the White student group. The classifiers were trained on sets that were dominated by the White student group and had very little representation from minority groups. This will obviously skew the accuracy and dependability of the model especially if it is tested on high numbers of minority groups.

1. Lessons Learned

Prior to this class, I had never learned Python and did not even know what machine learning was. I never knew how artificially intelligent machines were made or how they got smarter. This class has taught me so much. I wish I had taken an introduction course before this class because I think I would have been able to dig deeper into the material and understand the material on a deeper level. I learned how to clean up data using Python, create graphs in Python, maneuver and draw conclusions from graphs, learned how to train classifiers by method of logistic regression and random forest. I also learned how to create an explanation method using LIME.

1. References
2. https://contexts.org/articles/gender-in-the-one-percent/
3. [https://christophm.github.io](https://christophm.github.io/)
4. <https://www.nationsreportcard.gov/profiles/stateprofile?chort=1&sub=MAT&sj=&sfj=NP&st=MN&year=2019R3>
5. https://towardsdatascience.com/understanding-feature-engineering-part-2-categorical-data-f54324193e63
6. https://github.com/biplav-s/course-tai/blob/main/sample-code/l9-explanations/LIME%20explanations%20on%20tabular%20data.ipynb