Predictive Modeling and Fairness in Higher Education: A Case Study with FIU Admissions Data

Osmel Cereijo and Daniela Zaccardi
Knight Foundation School of Computing and Information Science
Florida International University
Course Instructor: Dr. Ananda M. Mondal
Mentor: Dr. Agoritsa Polyzou

Project Repository:
https://github.com/DanielaZaccardi/capstone_predictive_fairness

***Abstract*—Predictive models are increasingly used in university admissions, offering opportunities to improve decision-making but raising concerns about the amplification of historical inequities. This project investigates predictive modeling and algorithmic fairness using Florida International University's Fall 2024 undergraduate applicant dataset. After data preparation, including handling missing values, creating unified test score features, and discretizing key variables, we evaluated several machine learning models to assess predictive performance and identify the strongest predictors of admission. Results show that academic preparation and admission type are the most influential predictors, while demographic attributes do not significantly affect model decisions, an encouraging sign for fairness. However, fairness analysis across demographic groups revealed consistent disparities in selection rates and error-rate patterns, indicating structural imbalances in the underlying data rather than model-specific bias. To address these disparities, we applied pre-processing mitigation techniques, including Random Oversampling and SMOTE. Oversampling notably improved fairness for underrepresented groups with minimal impact on model performance. The study demonstrates that fairness can be enhanced without sacrificing accuracy and highlights the need for ongoing evaluation, broader datasets, and intersectional analyses in future work.**

***Index Terms*—predictive modeling, algorithmic fairness, admissions, machine learning classification, bias mitigation**

I. Introduction

A. Background

Universities increasingly rely on predictive analytics to support admissions decisions, offering consistency and efficiency in evaluating large applicant pools. However, these models may unintentionally replicate or amplify historical patterns of inequality embedded in the data. Bias in this context refers broadly to systematic disparities that arise when certain groups are overrepresented, underrepresented, or treated differently due to patterns learned from historical information [1,2].

Two forms of bias are particularly relevant in admissions. Machine learning bias emerges when model predictions disproportionately favor or disadvantage specific groups because of inequitable patterns in the training data. Statistical bias, on the other hand, reflects systematic errors in estimating outcomes or representing populations, such as inaccurate selection rates or misaligned group distributions. Both forms of bias can influence admissions outcomes by reinforcing existing disparities related to academic preparation, socioeconomic context, or demographic characteristics [3,4].

Ensuring that predictive systems operate fairly is critical, as admissions decisions profoundly shape students' educational and socioeconomic trajectories. Evaluating whether models introduce or exacerbate disparities, and understanding why those disparities occur, is essential for promoting transparency, accountability, and the ethical use of machine learning in higher education.

B. Goals and Objectives

The project is guided by two overarching goals, consistent with prior work emphasizing the need to understand and mitigate algorithmic bias in educational settings. The first goal is to explore how machine learning models can develop bias across different applicant groups. The second goal is to investigate methods to alleviate such bias while maintaining predictive accuracy, ensuring that fairness interventions do not compromise the utility of the predictive models.

Building from these goals, the project focuses on three core objectives. First, we aim to predict admissions outcomes using FIU's Fall 2024 undergraduate applicant dataset and identify which features contribute most strongly to model predictions. Second, we evaluate fairness across different demographic groups to determine whether models exhibit disparate outcomes or measurable bias. Finally, we investigate which bias mitigation methods are most effective at reducing disparities while retaining strong predictive accuracy, ensuring that improvements in fairness do not compromise model performance.

These objectives directly align with the project's research questions, which guide the analysis:

- How accurately can we predict admissions decisions?
- Which features most strongly influence predictions?
- Does the model exhibit bias, and if so, which groups are most affected?
- If bias exists, is it harmful or justifiable?

- What methods can detect, measure, and mitigate bias?
- How much bias can be effectively reduced without sacrificing accuracy?

Together, these goals, objectives, and research questions provide a structured framework for evaluating both the predictive and fairness dimensions of admissions modeling.

C. Motivation

Fairness in admissions has become increasingly important due to growing public scrutiny and evolving policy expectations. Federal actions, such as the U.S. Department of Education's audit of university admissions data and the heightened attention following the Students for Fair Admissions v. Harvard decision, reflect national concern about equitable access to higher education [5,6]. Institutions are expected to demonstrate transparency, accountability, and fairness in how they evaluate applicants. At the same time, disparities in educational experiences, such as differences in high-school quality, access to standardized test preparation, and socioeconomic resources, may be embedded in the data used to train predictive models. Without careful analysis, these underlying inequities can be reproduced by machine learning systems, potentially disadvantaging groups that are already marginalized [2,7].

Motivated by these legal, societal, and ethical pressures, this project seeks to examine both model performance and fairness using a real and recent university dataset. The goal is to provide evidence-based insight into whether predictive models used in admissions make equitable decisions and how fairness can be improved without compromising accuracy, ensuring that data-driven systems promote rather than hinder equal opportunity.

II. Literature Review

Research on algorithmic decision-making in education has expanded rapidly in recent years, driven by growing interest in how predictive models behave when applied to student data. Scholars have examined how different modeling approaches perform on academic, behavioral, and admissions-related tasks, highlighting both the potential of machine learning to support institutional decision-making and the risks associated with unequal error rates and disparate impacts across demographic groups. Much of this work focuses on understanding how predictive patterns emerge from historical inequalities and how these patterns shape model performance in high-stakes contexts [3,4]. Within this broader landscape, admissions prediction has become a focal domain for studying fairness in educational AI because it offers a clear opportunity to analyze group-level disparities and evaluate trade-offs between accuracy, equality, and transparency [8,9].

Across recent studies, several modeling approaches including logistic regression, decision trees, random forests, and various ensemble techniques, have been shown to perform effectively in predicting admission likelihood and other student outcomes. For example, Raftopoulos et al. (2024) demonstrates that these traditional ML models can achieve strong predictive accuracy across multiple educational datasets, often outperforming rule-based or manual selection processes. However, predictive accuracy alone is insufficient for evaluating whether models behave equitably, especially when historical data encode disparities in academic preparation or opportunity [8].

Multiple studies highlight that machine learning systems can inadvertently reproduce structural inequities present in educational contexts. Bird et al. (2024) shows that even when models do not explicitly use sensitive attributes such as race, they may still generate unequal error rates because correlated features encode underlying social patterns. Gandara et al. (2024) further demonstrates that racialized and marginalized groups often experience lower model precision or recall, underscoring the importance of examining subgroup performance when evaluating fairness. These findings suggest that fairness must be assessed at the group level and not solely through overall accuracy metrics. Prior work consistently stresses the importance of examining baseline group disparities before building predictive models, as underlying differences in representation, preparation, or opportunity can heavily influence both accuracy and fairness outcomes [3,4,8].

To support fairness evaluation, researchers increasingly rely on group fairness metrics such as statistical parity, disparate impact, equal opportunity, and equalized odds. Raftopoulos et al. (2024) emphasize that different fairness metrics capture distinct aspects of model behavior, and it is common for a model to satisfy one criterion while violating another. This complexity highlights the need for multi-metric fairness evaluation and for interpreting results within the specific decision-making context. The same study argues that transparent, interpretable models are essential for identifying whether sensitive attributes, or heavily correlated features, influence predictions [8].

Bias mitigation strategies have also received growing attention. Van Busum & Fang (2025) finds that dataset imbalances are a major contributor to disparities in admissions modeling, particularly when certain demographic groups are underrepresented in the training data. Their human-in-the-loop framework demonstrates that mitigation strategies often involve iterative adjustment of sampling or fairness constraints and must be tailored to institutional priorities [9]. Pre-processing approaches, such as oversampling, reweighting, or generating synthetic instances, are frequently recommended because they adjust the dataset before modeling and can reduce disparities without requiring specialized algorithms. This aligns with broader findings from AI fairness literature, where pre-processing techniques have been shown to improve group-level fairness while maintaining reasonable predictive performance.

Taken together, prior research highlights three interconnected insights. First, predictive models in education can offer strong performance but may exhibit uneven outcomes across demographic groups. Second, fairness evaluation requires a combination of group-level metrics and model interpretability techniques to understand both the magnitude and drivers of disparities. Third, mitigation often involves balancing com-

peting goals (i.e., fairness, accuracy, and transparency) since no approach eliminates bias universally.

III. Challenges

The development of fair and reliable predictive models for university admissions requires addressing several interconnected challenges. These challenges arise from both the structure of the dataset and the complexities of fairness evaluation, as noted across recent educational and fairness research [3,4,8,9].

A. Data Quality and Pre-Processing

Fairness evaluation is complicated by the fact that metrics such as statistical parity, disparate impact, equal opportunity, and equalized odds capture different and sometimes conflicting dimensions of fairness. Determining which metrics are most appropriate for admissions requires understanding both their technical meaning and their policy implications.

B. Bias Detection and Measurement

Improving model fairness often introduces trade-offs with predictive performance. Techniques like reweighting, resampling, or synthetic data generation can reduce disparities but may also cause overfitting or distort the data distribution. Achieving fairness while maintaining accuracy remains a central challenge in high-stakes decision contexts.

C. Bias Mitigation Trade-Offs

Reducing disparities across demographic groups often comes with the risk of degrading model performance. Mitigation techniques (e.g., reweighting, resampling, generating synthetic samples) can improve fairness but may introduce overfitting, distort the underlying data distribution, or weaken predictive accuracy. Balancing fairness improvements with performance retention is a central challenge when deploying models in high-stakes decision-making environments.

D. Model Transparency and Explainability

Meaningful fairness assessment requires models that are sufficiently interpretable to reveal whether sensitive attributes or correlated variables influence predictions. Tools such as feature importance analysis, rule-based models, and interpretable ML frameworks help identify potential sources of bias and support clear communication of results to institutional stakeholders who must ensure accountability.

IV. Data Sources and Description

We used an anonymized dataset provided by FIU containing undergraduate admissions applicants for the Fall 2024 semester. Before requesting the dataset, we reviewed a sample undergraduate application to understand the types of information typically collected during this process. This exercise helped us identify broad categories of potentially relevant variables, such as demographic characteristics, academic preparation, and application-related factors, and ensured that our data request aligned with institutional practices.

The original file shared included 48,426 undergraduate and graduate applicants. However, because the scope of this project is undergraduate admissions, all graduate applicants were removed. This resulted in 29,829 undergraduate records for analysis and 42 features, grouped into the following categories:

- Demographic Information (6): gender, age, ethnicity, country of birth, U.S. military status, and Florida residency.
- High-School and Academic Background (6): highest level of education, high-school state, type and GPA, undergraduate GPA, and graduate GPA.
- Standardized Test Scores (22): ACT composite score, ACT sub-scores (English, Math, Reading, Science, Writing), ACT-to-SAT conversions (composite and English/Reading/Writing sub-scores), SAT old scores (composite and Math/Verbal/Writing sub-scores), SAT new scores (composite and Math/Reading/Writing sub-scores), and SAT retake indicator.
- College Application Details (8): admitted, admission type, intended major, Honors College admission, denied, incomplete, matriculated, and enrolled.

A. Data Quality Issues and Challenges

1) Data Quality: First, we identified inconsistencies in the admitted and denied features. Specifically, 3,457 applicants were labeled as neither admitted nor denied. To avoid losing an additional 12% of the data, we treated these cases as not admitted for the purposes of our analysis. Additionally, we found that the Ethnicity feature contained a "non-resident alien" category. This classification does not represent an actual ethnicity, and there were no other variables in the dataset that allowed us to reliably reassign this group. As a result, we retained the category as provided.

2) High Cardinality Features: Two variables contained a large number of unique categories: intended major and country of birth. To reduce dimensionality and improve model interpretability, we consolidated values in both features. For intended major, we mapped all programs to their corresponding FIU colleges, reducing the number of categories from 134 to 9. Similarly, for country of birth, we grouped countries into their respective continents, decreasing the number of categories from 196 to 7.

3) Missing Values: A substantial portion of the dataset contained missing values, particularly across standardized test scores and cumulative GPA scores. Overall, 26 of the 42 features exhibited more than 50% missingness. Thirteen features had 90% or more missing values, including undergraduate and graduate cumulative GPA (100% missing), SAT (old format) composite and sub-scores (99%–100% missing), and ACT composite and sub-scores (92%–99% missing). An additional thirteen features showed 50%–90% missingness, such as ACT-to-SAT conversion scores (86%), SAT (new format) composite and sub-scores (59%–64%), and high-school type and state (51%–53%). To improve data usability and reduce sparsity in the standardized testing variables, we consolidated all available test information into a single unified SAT-based score. This

was achieved by combining reported SAT total with ACT-to-SAT conversion scores present in the dataset. Consolidating the standardized scores allowed us to retain academic performance information for a larger portion of applicants, reduce overall missingness to roughly 50%, and create a consistent, continuous feature suitable for modeling.

At this point, any feature with more than 50% missing values was excluded from further analysis. For categorical variables, we also standardized labels reflecting "missing" or "not reported" to ensure consistency. In addition, we created binary indicator variables for missing high-school GPA and SAT unified score to aid in our analysis.

After the cleaning process, we retained most of the original feature-mix in a more structured and usable form. The final set of features to conduct exploratory data analysis (EDA) and bias investigation included:

- Demographic Information (6): gender, age, ethnicity, continent of birth, U.S. military status, and Florida residency.
- High-School and Academic Background (3): highest level of education, high-school GPA, and high-school GPA missing (flag).
- Standardized Test Scores (2): SAT total unified score and SAT total unified score missing (flag).
- College Application Details (8): admitted, admission type, and intended college.

B. Data Overview

Before conducting EDA, we examined the overall characteristics of the applicant population.

1) Admission Rate: Approximately 48% of applicants were admitted, indicating a moderately selective process.

2) Applicant Profile: Most applicants (71%) were first-time-in-college (FTIC), consistent with FIU's undergrad freshman profile. A majority (58%) identified as female, and 57% were Florida residents, which aligns with FIU's status as a large public institution serving the state's population.

3) Age Distribution: Applicant ages ranged from 15 to 68, though the distribution is highly right-skewed as shown in Fig. 1. Most applicants fall within a narrow traditional college-entry range: 75% are age 20 or younger, and the median age is 18 (mean = 20, SD = 4.3). A smaller number of older or returning adult learners create a long positive tail.

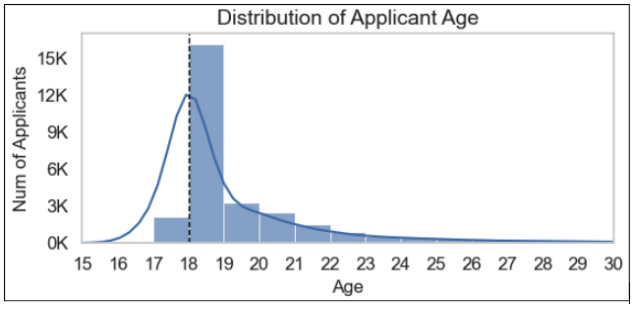

Fig. 1. Histogram of applicant population age (x-axis truncated at 30).

4) Ethnicity: The applicant pool is ethnically diverse, with proportions consistent with FIU's demographic composition as a Hispanic-Serving Institution. As expected, Hispanic or Latino applicants constituted the largest group (39%), followed by Non-resident Alien applicants (22%). White (16%) and Black or African American applicants (15%) represented comparable shares of the population. The remaining 8% of applicants were grouped into an "Other or Unknown" category, which includes applicants reporting Two or More Races, Asian, Pacific Islander, and American Indian or Alaska Native.

5) Highest Level of Education: Most applicants (82%) reported a high-school diploma or GED, while 15% reported an associate degree, and a small percentage reported higher levels. This distribution is expected given the high share of FTIC applicants.

V. Exploratory Data Analysis and Bias Investigation

A. Visual Patters

Before developing predictive models, we conducted a detailed exploratory data analysis (EDA) to better understand the structure, distribution, and relationships within the dataset. This step is essential for identifying underlying patterns, detecting potential sources of bias, and ensuring that the modeling process is informed by the characteristics of the applicant population.

1) Gender: Female applicants constitute most of the applicant pool, representing 57.5% of all submissions, compared to 42.5% male applicants. Despite this difference in application volume, the admission outcomes are relatively similar across genders as shown in Fig 2. Female applicants were admitted at a rate of 46.8%, while male applicants had a slightly higher admission rate of 49.3%. These differences are modest and do not indicate substantial gender-driven disparities in admissions decisions. Because the proportion of female applicants is notably higher than that of male applicants, the overall admitted pool also reflects this skewed distribution even though the admission rates themselves are comparable. This pattern suggests that gender does not appear to be a significant driver of admissions outcomes in this dataset.

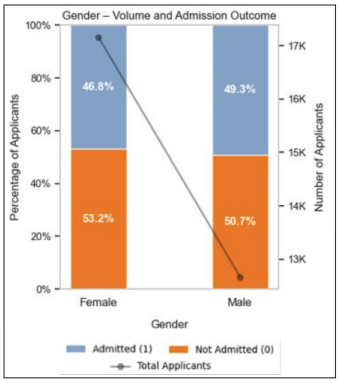

Fig. 2. Gender admission outcomes, with applicant volume shown on the secondary axis.

2) Florida Residency: Florida residency shows a clear divide in both applicant volume and admissions outcomes. In-state students make up 56.9% of the applicant pool, while 43.1% apply from out of state. The admission patterns differ substantially between the two groups as seen in Fig 3. In-state applicants experience a significantly higher admission rate (66.3%) compared to out-of-state applicants (23.5%), indicating that residency status plays an important role in FIU's admissions decisions. This disparity is consistent with enrollment practices at large public institutions, where the number of out-of-state applicants who can be admitted is often constrained by institutional or state caps. As a result, out-of-state applicants represent a smaller segment of the applicant pool, with a far more selective process. These differences highlight the importance of including residency status as a key feature in predictive modeling and fairness analysis, given its strong and policy-driven influence on admission likelihood.

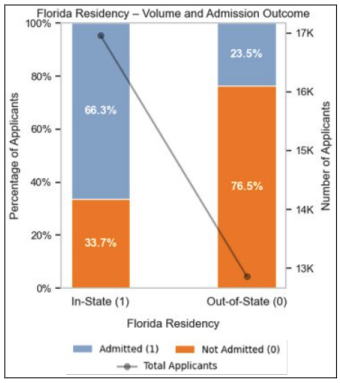

Fig. 3. Florida residency admission outcomes, with applicant volume shown on the secondary axis.

3) Military Status: Military status represents a very small portion of the applicant population, with 97% of applicants reporting no military affiliation and only 3% identifying as dependents, veterans, active-duty personnel, reservists, or National Guard members combined. Despite the small subgroup size, applicants with a military affiliation show a slightly higher admission rate (54.2%) compared to non-military applicants (47.7%) as show in Fig 4. While this suggests a potential institutional or policy consideration for military-connected students, the extremely small representation of this group limits the reliability of these estimates. Such small sample sizes can lead to unstable statistical patterns, increased variance in model predictions, and challenges in fairness evaluation. Therefore, although military status is an important demographic characteristic to monitor, its limited presence in the dataset warrants caution when interpreting group-level trends or incorporating this feature into predictive modeling.

4) Ethnicity: Ethnicity exhibits notable differences in both applicant distribution and admission outcomes. Hispanic/Latino students constitute the largest group in the applicant pool (39%), reflecting FIU's status as a Hispanic-Serving Institution. Non-resident Alien applicants represent the second-largest group (22%), followed by White (16%) and

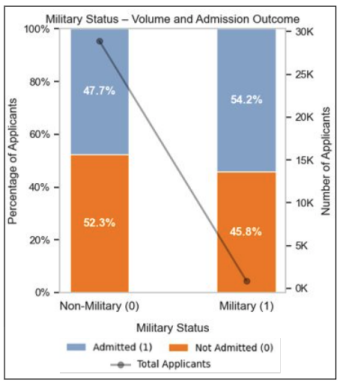

Fig. 4. Military status admission outcomes, with applicant volume shown on the secondary axis.

Black or African American applicants (15%), while 8% fall into the Other/Unknown category. Admission patterns vary widely across groups. Hispanic/Latino (60.1%) and Asian applicants (63.4%) experience the highest admission rates, whereas Black or African American (29.5%) and Non-resident Alien applicants (36.0%) face substantially lower rates, indicating more selective outcomes for these groups. White applicants fall near the middle with a 49.1% admission rate. These disparate patterns (Fig 5) highlight the importance of examining ethnicity as a key dimension of fairness, as certain groups experience disproportionately lower admission rates relative to their representation in the applicant pool.

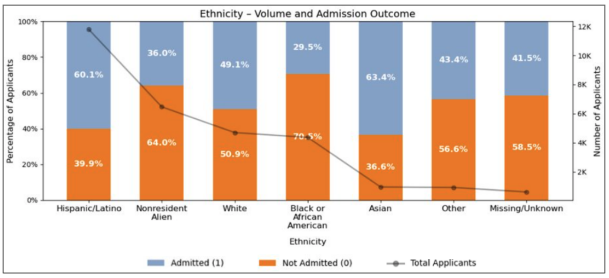

Fig. 5. Ethnicity admission outcomes, with applicant volume shown on the secondary axis.

5) Age: The age distribution of applicants is heavily concentrated among traditional college-age students. Most applicants are between 17 and 20 years old, with a median age of 18 and a mean age of 20. Approximately 75% of all applicants are 20 years old or younger, while older applicants make up only a small share of the population, forming a long but thin right-tail in the distribution. This pattern is consistent with expectations for a four-year public university serving primarily FTIC students.

Admission outcomes, however, show notable differences across age groups (Fig 6). Applicants under age 18 had the lowest admission rate in the dataset (25.1%), likely reflecting stricter evaluation. Admission rates improve sharply for 18 and 19-year-old applicants, who experience selection rates between 39% and 42%. The highest admission rates occur

among applicants aged 20 to 30, ranging from 68% to 71%, indicating that slightly older applicants, many of whom may be transfer, returning, or non-traditional students, have a stronger likelihood of acceptance. Admission rates decline again past age 30 (66.8%), though this group makes up only a small fraction of total applicants.

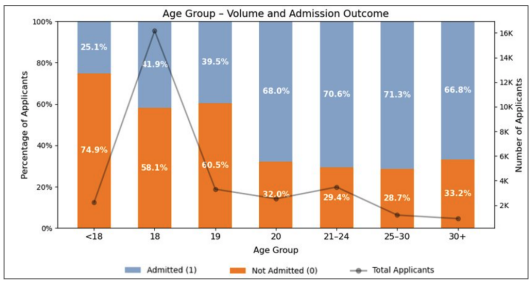

Fig. 6. Age admission outcomes, with applicant volume shown on the secondary axis.

6) Continent of Birth: The distribution of applicants by continent of birth is highly uneven, with most born in North America (77.4%), followed by smaller proportions from South America (7.6%), Asia (7.3%), Europe (2.9%), and Africa (2.7%), with the remaining applicants falling into the Oceania or Missing/Unknown categories.

Admission patterns vary substantially across continents as shown in Fig 7. Applicants born in South America have the highest admission rate (60.5%), followed by those from Europe (46.0%) and North America (49.6%). In contrast, applicants from Asia experience a markedly lower admission rate (29.7%), and applicants born in Africa have the lowest admission rate in the dataset (13.3%). These results highlight pronounced disparities in admission likelihood across birth regions, which may partially reflect structural factors such as Florida residency policies, international applicant caps, or other aspects of the admissions process that differ for foreign students.

Notably, these findings differ from the patterns observed in the ethnicity analysis. While Asian applicants (based on self-reported ethnicity) exhibited one of the highest admission rates (63.4%), applicants born in Asia had one of the lowest (29.7%). This contrast suggests that continent of birth and ethnic identity capture different dimensions of applicant background and may interact differently with institutional admissions criteria. It underscores the importance of analyzing multiple fairness-related attributes independently rather than assuming consistent trends across related variables.

7) College of Intended Major: Although intended major is not a demographic characteristic, it plays a meaningful role in admissions outcomes and is important to evaluate due to its potential influence on both predictive modeling and fairness assessments. Applicant volume varies widely across colleges, with the largest shares applying to Arts, Sciences & Education (27.2%), Business (23.5%), and Engineering & Computing (18.1%). Admission rates also differ substantially across colleges, reflecting variations in program capacity,

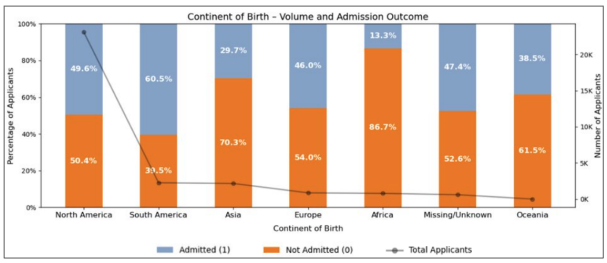

Fig. 7. Continent of birth admission outcomes, with applicant volume shown on the secondary axis.

competitiveness, and institutional priorities. Nursing & Health Sciences exhibits the lowest admission rate at 22.6%, followed by Public Affairs at 37.8%. In contrast, Arts, Sciences & Education and Engineering & Computing show higher admission rates (57.3% and 53.6%, respectively), even at relatively high application volumes. These patterns (summarized in Fig 8) demonstrate that intended major can be an influential predictor of admission likelihood and should be included in modeling to capture program-level effects.

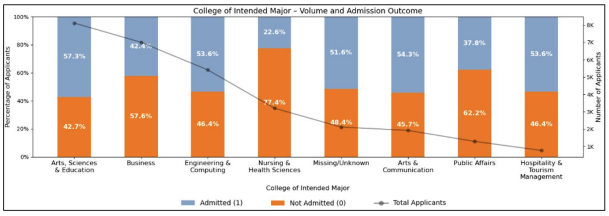

Fig. 8. College of intended major admission outcomes, with applicant volume shown on the secondary axis.

8) Application Type: Application type shows some of the strongest differences in admission outcomes across the entire applicant population (Fig 9). Most applicants (70.5%) are FTIC students, yet this group experiences the lowest admission rate at 36.8%. In contrast, all other admission categories show significantly higher acceptance rates. Community College (CC) Transfer applicants, who make up 15.6% of the pool, are admitted at a much higher rate (86.4%), and Other Transfer students (11.3% of the pool) also see more favorable outcomes (53.6% admitted). The most selective advantage appears among Second Bachelor's applicants and Undergraduate Non-FTIC applicants, who have exceptionally high admission rates (88.9% and 96.3%, respectively), though these groups represent less than 3% of the total applicants.

These patterns indicate that FTIC admissions are far more competitive than transfer or post-baccalaureate pathways, which is typical for large public institutions where freshman admission slots are capped and transfers are often admitted through separate processes. Because FTIC students represent such a large proportion of the pool and face substantially lower acceptance rates, this feature will likely play an influential role in predictive modeling.

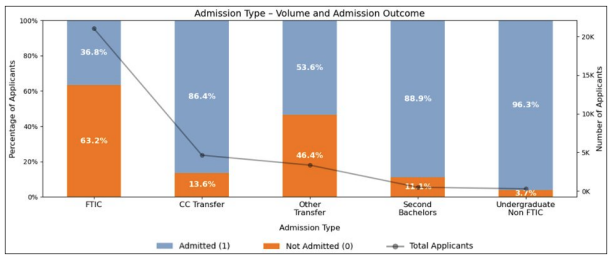

Fig. 9. Application type admission outcomes, with applicant volume shown on the secondary axis.

B. Quantitative Metrics

The visual EDA reveals clear differences in admission outcomes across groups, but visual trends alone cannot confirm or quantify bias. To evaluate these disparities more rigorously, we leverage selection rate, weighted selection rate, and entropy difference to measure group-level gaps in a consistent and interpretable way for demographic attributes [2,8,10].

1) Selection Rate Difference: Selection rate (SR) measures the proportion of applicants in a given group i who are admitted as shown in 1. In the context of admissions, SR represents the base measure for comparing outcomes across demographic groups. Meaningful differences in SR may indicate unequal treatment or potential bias, particularly when one group experiences systematically higher or lower admission rates.

$$SR_i = \frac{Admitted_i}{Applicants_i} \tag{1}$$

To quantify disparities between groups, we compute the Selection Rate Difference (SRD), defined in 2 as the gap between the highest and lowest SR across all groups for a given attribute. For features with only two groups, SRD reduces to a simple difference between the two rates. For attributes with multiple categories (i.e., Ethnicity) the metric captures the range between the maximum and minimum SR values. A value close to zero indicates parity, while larger positive or negative differences signal unequal likelihood of admission across groups [2,8,10].

$$SRD = max(SR_i) - min(SR_i) \tag{2}$$

2) Weighted Selection Rate Difference: Since demographic groups vary widely in size, the weighted selection rate (WSR) as defined in equation 3 expands on SR by incorporating each group's weight (equation 4) into the calculation. This adjustment ensures that groups contribute to the overall metric in relation to their actual representation, rather than being treated equally regardless of size.

$$WSR_i = w_i * SR_i = \frac{Admitted_i}{\sum_j \text{Applicants}_j} \tag{3}$$

$$w_i = \frac{Applicants_i}{\sum_j \text{Applicants}_j} \tag{4}$$

The weighted selection rate difference (WSRD) measures the gap between the groups that contribute the most and least to the admission total as defined in equation 5. Ideally, this difference should be as close to zero as possible, indicating that no group disproportionately influences the admission total [2,8,10].

$$WSRD = max(WSR_i) - min(WSR_i) \tag{5}$$

3) Entropy Difference: Entropy measures the diversity or balance of applicants across groups and provides a distribution-level perspective on fairness. In this context, entropy captures how evenly demographic groups are represented within the applicant pool versus the admitted pool. Higher entropy indicates greater diversity and a more balanced distribution across categories, while lower entropy indicates that one or more groups dominate the composition. We calculate entropy (H) separately for the applicant pool (equation 6) and the admitted pool (equation 7).

$$H(p) = -\sum_i p_i \log_2(p_i) \tag{6}$$

$$H(q) = -\sum_i q_i \log_2(q_i) \tag{7}$$

where

$$p_i = \frac{\text{Applicants}_i}{\sum_j \text{Applicants}_j}, \qquad q_i = \frac{\text{Admitted}_i}{\sum_j \text{Admitted}_j} \tag{8}$$

The entropy difference (ED), defined in equation 9, measures the shift in diversity between these two distributions. This metric evaluates fairness by assessing whether the admission process preserves the diversity present in the applicant population. A small entropy difference indicates that the demographic balance of admitted students closely reflects that of the full applicant group. Conversely, a larger entropy difference signals that certain groups become disproportionately under or over-represented after admissions decisions are made. Unlike SR-based metrics, entropy captures multi-group imbalance simultaneously and is particularly useful for identifying structural skew that may not be visible when examining two groups at a time [2,8,10].

$$\Delta(H) = H(p) - H(q) \tag{9}$$

4) Interpretation of Group-Level Metric Results: Table I summarizes how each demographic feature groups differ across three metrics. Overall, most demographic attributes exhibit non-zero disparities, indicating that admissions outcomes are not evenly distributed across groups.

TABLE I
SUMMARY OF DIFFERENCES ACROSS DEMOGRAPHIC FEATURES

Feature	SRD	WSRD	ED
Gender	0.025	0.060	0.005
Florida Residency	0.428	0.276	-0.241
US Military Status	0.066	0.446	0.020
Ethnicity	0.338	0.229	-0.154
Age	0.463	0.208	0.138
Continent of Birth	0.472	0.384	-0.155

Gender shows the smallest differences across all three metrics, suggesting that admissions outcomes are relatively balanced between male and female applicants. Florida Residency displays the largest disparities, with substantial gaps in both selection rates and contribution weights. This reflects FIU's prioritization of in-state applicants and the structural enrollment limits imposed on out-of-state admissions. Ethnicity and Continent of Birth also exhibit moderate disparities, particularly through negative entropy differences that indicate reduced diversity in admitted students relative to the applicant pool. U.S. Military Status shows relatively small disparities, however, reliability might be impacted due to the subgroup's size. Age presents a modest positive entropy difference, suggesting slightly greater diversity among admitted applicants. Taken together, these patterns highlight which demographic characteristics are more strongly associated with differential admissions outcomes and where representation shifts the most between applicants and admitted students.

VI. Methodology

A. Experimental Set-Up

The experimental workflow for this study follows a structured pipeline designed to ensure methodological consistency, fairness evaluation, and reliable model comparison. Figure 10 summarizes the key stages, beginning with an 80/20 stratified train–test split to preserve the original admission ratio. The training portion undergoes pre-processing, including k-means discretization for continuous features and one-hot encoding for categorical variables. Models are then tuned through 5-fold cross-validation, retrained using the optimal hyper-parameters, and finally evaluated on the held-out test set.

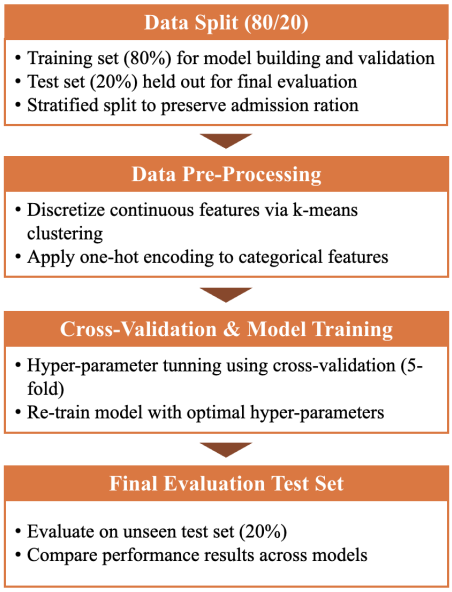

Fig. 10. Summary of Experimental Set-Up Pipeline

1) Feature Discretization: Continuous variables (e.g., age, high-school GPA, SAT total score) were transformed into categorical bins using k-means clustering to capture natural groupings in the data rather than relying on arbitrary cut-points. The number of clusters was selected using the elbow method, which identifies the point where additional clusters provide diminishing reductions in within-cluster variance. To avoid data leakage, the k-means models were fit exclusively on the training split, and the learned cluster boundaries were then applied to transform the test set.

For age, k-means with k=2 effectively separates younger and older applicants while avoiding overly small clusters. This split captures meaningful demographic variation as shown in Fig 11.

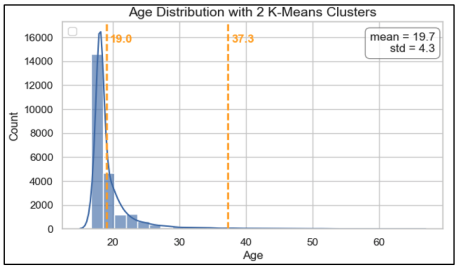

Fig. 11. Distribution of Applicant Age with K-Means Cluster Boundaries

For high-school GPA, the elbow method supports k=3, which also aligns with interpretable academic classification categories (low, mid, high GPA). These clusters maintain balanced group sizes and reflect meaningful academic differences among applicants as shown in Fig 12.

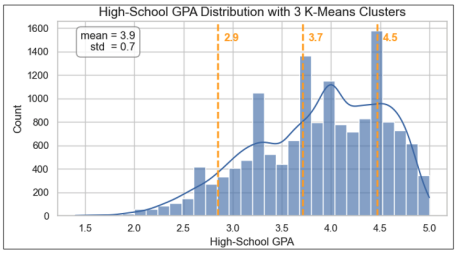

Fig. 12. Distribution of High-School GPA with K-Means Derived GPA Bands

The SAT unified score distribution is broad, slightly right-skewed, and exhibits multiple density peaks corresponding to performance bands as shown in Fig 13. K-means with k=3 captures these natural groupings well and corresponds to low, medium, and high academic performance levels.

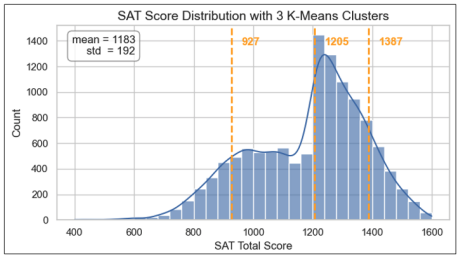

Fig. 13. Distribution of SAT Total Scores with K-Means Derived Score Bands

Table II summarizes the final discretization for all continuous variables. For each feature, the optimal number of clusters, the resulting value ranges, and their assigned category labels are reported.

TABLE II
SUMMARY OF DISCRETIZED CONTINUOUS FEATURES

Feature	Clusters	Range	Label
Age	2	15–28	Younger Applicants
		29–68	Older Applicants
High-School GPA	3	1.4–3.2	Low GPA
		3.3–4.0	Mid GPA
		4.1–5.0	High GPA
SAT Total Score	3	400–1060	Low Score
		1070–1290	Mid Score
		1300–1600	High Score

Note: An extra cluster code (–1) was assigned to identify missing values for High-School GPA and SAT Total Score.

B. Modeling Approaches

This study evaluates several machine-learning models that represent diverse learning paradigms (e.g., tree-based, linear, probabilistic, and ensemble methods) to compare predictive performance and select a baseline model to further examine for fairness.

1) Decision Trees: Hierarchical, rule-based classifiers that recursively split the data using the features that provide the greatest reduction in impurity. They are particularly useful in educational admissions settings because they produce transparent and interpretable decision rules. Advantages also include their ability to handle mixed data types, operate without feature scaling, and inherently manage missing values. However, their main drawbacks are their tendency to overfit when not properly pruned and their potential bias toward features with many categories, which can lead to instability across different training samples [11,12].

2) Random Forest: Ensemble learning method that constructs many decision trees using bootstrapped samples and random subsets of features. This approach improves predictive robustness and reduces overfitting compared to a single decision tree. Key strengths of this method also include strong performance on tabular data, built-in estimates of feature importance, and its resilience to noise. On the downside, the ensemble structure makes the model slightly less interpretable, and it requires greater computational resources during both training and inference. These properties make it well-suited for high-dimensional admissions data [11,12].

3) AdaBoost (Adaptive Boosting): Sequential ensemble technique where each new weak learner, typically a shallow decision tree, focuses on correcting the classification errors of previous learners. The method often yields higher accuracy than single-tree models and performs well with weak, low-complexity learners. However, AdaBoost is highly sensitive to noisy data and outliers, and excessive boosting rounds can cause overfitting [11,12]. Despite these limitations, AdaBoost offers a strong balance between performance and complexity for admissions prediction tasks.

4) Logistic Regression: Linear probabilistic model that estimates the likelihood of admission using a sigmoid function. It is valued for its interpretability, as coefficients directly indicate the direction and strength of feature effects. Logistic regression is computationally efficient, requires minimal tuning, and makes no assumptions about feature distributions. However, it assumes linear relationships, which may lead to under-performance when interactions or nonlinear patterns are present [11,12]. In fairness contexts, its transparency is an advantage, but its simplicity can miss structural patterns in admissions data.

5) Gaussian Naïve Baye: Probabilistic classifier that assumes features follow a normal distribution and contribute independently to the prediction. The model trains extremely quickly and performs well on smaller datasets or scenarios where independence assumptions approximately hold. Yet, the strict assumption of independence between features and sensitivity to distributional mismatches can degrade performance when features interact or exhibit skewed distributions [11,12]. Despite this, its efficiency and simplicity make it a strong model for comparison.

6) Bernoulli Naïve Bayes: Variant designed for binary features, such as one-hot encoded categorical variables used in this project. It is efficient, simple to implement, and performs well with sparse data. However, like other Naïve Bayes variants, it assumes feature independence and struggles when important interactions occur among features [11,12]. Although not always the most accurate classifier, it serves as a fast benchmark and aligns with the feature types in our preprocessed dataset.

To ensure fair comparison across models, all classifiers were tuned using grid search with 5-fold stratified cross-validation. The hyperparameters were selected based on standard machine learning literature and practical recommendations for each algorithm [11,12]. For Decision Trees and Random Forests, we varied tree depth, minimum samples per split and leaf, impurity criteria, and feature-sub-sampling strategies. Logistic Regression was tuned across L1, L2, and no-penalty configurations with a range of regularization strengths and compatible solvers. For AdaBoost, we explored different numbers of estimators and learning rates to balance bias–variance trade-offs.

C. Fairness Evaluation Metrics

Fairness assessment requires comparing model behavior across different demographic subgroups categorized as protected (or unprivileged) and unprotected (or privileged) groups. Protected groups are those that may face structural disadvantages, historical under-representation or differential treatment, and are therefore monitored to ensure the model does not disproportionately reduce their chances of a favorable outcome (e.g., admission). In contrast, unprotected groups tend to receive more favorable outcomes and serve as the reference when computing fairness metrics [2].

In our admissions dataset, protected groups are defined as those with lower selection (admission) rates, such as females and out-of-state applicants, while unprotected or privileged groups, such as males and in-state applicants, exhibit higher selection rates. These group definitions guide the calculation of the following fairness metrics [2].

1) Statistical Parity or Demographic Parity: Evaluates whether the likelihood of receiving a positive prediction (e.g., admission) is the same across unprivileged and privileged groups as defined in equation 10.

$$\text{SP} = P(\hat{Y} = 1 \mid A = \text{unpriv}) \; - \; P(\hat{Y} = 1 \mid A = \text{priv}) \quad (10)$$

A value close to zero indicates that both groups are selected at similar rates. Negative values imply that the unprivileged group is receiving positive outcomes at a lower rate than the privileged group, signaling potential bias in the model or underlying data [2]. Statistical parity is widely used because it captures imbalance in outcomes regardless of underlying qualification differences.

2) Disparate Impact: Unlike statistical parity, disparate impact uses a ratio of selection rates [2]. It measures whether the rate of positive outcomes for a protected group is disproportionately lower than that of the privileged group as defined in equation 11. A value close to 1.0 indicates similar treatment between groups, while values below the commonly used 0.8 threshold (the "80% rule") suggest potential disparate impact against the protected group. This metric is particularly useful in admissions settings because it highlights whether protected groups receive successful outcomes at substantially lower rates relative to their peers.

$$\text{DI} = \frac{P(\hat{Y} = 1 \mid A = \text{unpriv})}{P(\hat{Y} = 1 \mid A = \text{priv})} \quad (11)$$

3) Equal Opportunity: Evaluates fairness by requiring that the true positive rate (TPR) be the same across unprivileged and privileged groups. In the context of admissions, this means that applicants who should be admitted (i.e., truly qualified) should have an equal chance of being predicted as admitted regardless of their demographic group. The metric is computed as the difference between the TPR of the unprivileged group and that of the privileged group as defined in equation 12. A value close to zero indicates fairness, while a negative value suggests that the protected group receives fewer TP predictions, signaling potential disparate treatment [2].

$$\text{EO} = \text{TPR}_{\text{unpriv}} - \text{TPR}_{\text{priv}} \quad (12)$$

where

$$\text{TPR} = \frac{TP}{TP + FN} \quad (13)$$

4) Equalized Odds: Stricter group-fairness metric that requires both true positive rates (TPR) and false positive rates (FPR) to be equal across unprivileged and privileged groups as defined in equation 14. While Equal Opportunity focuses only on equalizing TPRs, Equalized Odds enforces parity in both types of classification errors, ensuring that no group disproportionately benefits from correct predictions or suffers from incorrect ones [2]. A value close to zero indicates that the model treats both groups similarly in terms of TPR and FPR. Larger values indicate potential fairness concerns.

$$\text{EO} = |\text{TPR}_{\text{priv}} - \text{TPR}_{\text{unpriv}}| \; + \; |\text{FPR}_{\text{priv}} - \text{FPR}_{\text{unpriv}}| \quad (14)$$

where

$$\text{TPR} = \frac{TP}{TP + FN}, \qquad \text{FPR} = \frac{FP}{TN + FP} \quad (15)$$

D. Bias Mitigation

To address the disparities identified in our exploratory and quantitative analyses, we incorporated pre-processing bias mitigation techniques into the modeling pipeline.

E. Pre-Processing Bias Mitigation Techniques

These techniques aim to reduce bias before the model is trained by modifying the structure or distribution of the training data. These methods operate independently of the learning algorithm, making them widely applicable across model families. The primary objective is to increase the representation of protected or underprivileged groups such that the model can learn more balanced decision boundaries [2,13].

One of the most common pre-processing strategies is oversampling, which increases the number of training samples belonging to the unprivileged or underrepresented group. Two variants are typically used:

- Random Oversampling: Duplicates existing samples in the minority group to increase its prevalence in the training data. This method is simple and preserves the original data distribution but can increase the risk of overfitting due to repeated instances [14].
- SMOTE (Synthetic Minority Oversampling Technique): Generates synthetic samples by interpolating between observed minority-group instances. It reduces overfitting risks associated with simple duplication, but it may introduce noisy or less realistic samples depending on complexity of feature distributions [11].

By altering the group proportions in the training set only, the technique prevents leakage of synthetic or duplicated data into evaluation sets, maintaining a valid fairness assessment downstream [15]–[17].

F. Mitigation Procedure and Experimental Setup

1) Re-balance the Unprivileged Group: Oversampling was applied only to the training split to increase the representation of the unprivileged group among admitted applicants. We tested both Random Oversampling and SMOTE techniques, as well as two levels of adjustment for each:

1) Moderate adjustment that reduced the imbalance
2) Enforced equal representation (e.g., 50/50 distribution)

Evaluating these four configurations allowed us to isolate the effects of moderate vs. aggressive re-balancing and basic vs. synthetic augmentation within a consistent experimental framework.

2) Retrain Baseline Model: The baseline classifier (best performing model) was retrained separately for each oversampling configuration using the same hyper-parameters as the original model.

3) Evaluate on Test Set: Evaluate how each classifier performed based on each oversampling configuration. To preserve the integrity of the evaluation the test set was never oversampled or modified.

4) Recompute Performance and Fairness Metrics: After testing each model variant, we re-evaluated standard performance metrics for the overall model. Additionally, we looked at performance metrics and group fairness metrics for the specific attribute we over-sampled for.

VII. Results and Discussion

A. Baseline Model Performance

Across all six models evaluated, predictive performance was consistently strong, with accuracy, F1-score, and ROC-AUC values exceeding 0.850 on both validation and test sets (Table III). Among these, Random Forest, Decision Tree, and Logistic Regression emerged as the top-performing models.

Random Forest achieved the highest overall performance, with a test accuracy of 0.915, F1-score of 0.914, and ROC-AUC of 0.969, indicating strong separation between admitted and non-admitted applicants. Because it consistently outperformed other models across all major metrics, Random Forest was selected as the baseline model for all subsequent fairness evaluations and bias-mitigation experiments.

The Decision Tree model also performed competitively (test ROC-AUC 0.968), showing that even a single optimized tree can capture meaningful patterns in the admissions data. Logistic Regression likewise delivered strong results, achieving high recall (0.943) and a solid ROC-AUC (0.956). In comparison, the Naïve Bayes models showed lower precision and F1-scores, likely due to their strong independence assumptions, which may not hold in this dataset. AdaBoost provided moderate performance but did not surpass the ensemble strength of Random Forest.

Taken together, these findings indicate that tree-based models, Random Forest in particular, provide the most robust predictive foundation for the fairness analysis and bias-mitigation procedures that follow.

B. Feature Importance Analysis

Feature importance analyses were conducted using the Random Forest, Decision Tree, and Logistic Regression models, focusing on the top ten predictors identified by the Random Forest as shown in Fig 14. Tree-based models compute importance based on the total reduction in impurity (Gini or entropy) contributed by each feature across all splits [11,12]. The decision tree importance scores were based on reductions in entropy, while Random Forest used Gini impurity in the case of our implementation. Logistic Regression coefficients, were converted to standardized absolute values and normalized to sum to 1 so they could be compared directly with the tree-based importances.

Across all three models, academic and application-related variables consistently emerged as the strongest predictors of admission. In particular, Admission Type (FTIC), High-School GPA, and SAT Total Score dominated the importance rankings. These features were consistently assigned the highest weights in both the Random Forest and Decision Tree models, and the largest standardized coefficients in Logistic Regression. In contrast, demographic attributes were not among the top predictors, indicating that the models' predictive behavior is largely driven by academic qualifications and structural application features rather than sensitive characteristics.

Overall, the convergence of importance rankings across linear and non-linear models suggests that admission decisions in the dataset are primarily shaped by academic performance indicators, and that demographic variables play a comparatively minor role in the predictive structure. This is a favorable result from a fairness perspective, as it indicates that sensitive attributes were not major drivers of model prediction.

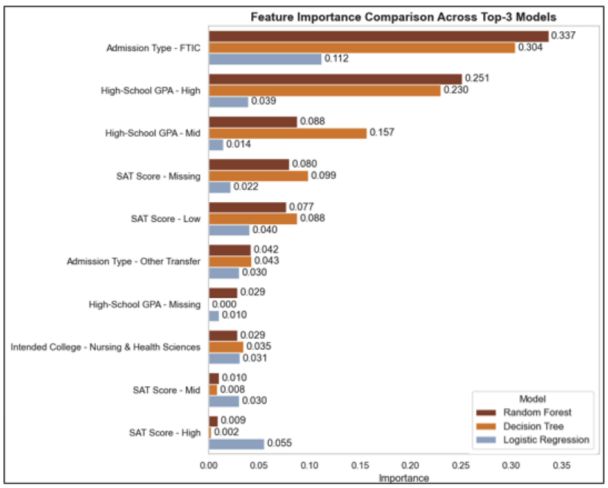

Fig. 14. Top 10 Features Identified by Random Forest (with Corresponding Importance from Decision Tree and Logistic Regression).

C. Fairness Evaluation Results

The fairness evaluation of the baseline Random Forest model reveals substantial differences across demographic groups when measured using the four standard group-fairness metrics summarized in table IV. As expected, some attributes exhibit minimal disparities while others exhibit more concerning gaps.

Gender shows very small differences across all metrics, suggesting that admissions outcomes are largely comparable. Military status also exhibits relatively minor disparities, though equalized odds (0.1917) suggests that error rates differ slightly between military and non-military applicants. However, military status was one of the smallest groups (3%) in our dataset, which can impact the accuracy or reliability of these results.

In contrast, Florida Residency shows the largest disparities across all metrics, confirming earlier exploratory findings. Out-of-state applicants have substantially lower selection rates

TABLE III
PERFORMANCE OF MACHINE LEARNING MODELS

Model	Accuracy		Precision		Recall		F1-Score		ROC-AUC	
	Val	Test	Val	Test	Val	Test	Val	Test	Val	Test
Gaussian Naïve Bayes	0.866	0.869	0.859	0.860	0.863	0.869	0.861	0.864	0.921	0.925
Bernoulli Naïve Bayes	0.853	0.853	0.810	0.807	0.904	0.910	0.855	0.856	0.926	0.927
Logistic Regression	0.912	0.911	0.883	0.879	0.940	0.943	0.911	0.910	0.960	0.956
Decision Tree	0.916	0.914	0.891	0.884	0.941	0.943	0.915	0.913	0.971	0.968
Random Forest	0.917	**0.915**	0.890	**0.886**	0.942	**0.944**	0.915	**0.914**	0.972	**0.969**
AdaBoost	0.898	0.895	0.888	**0.886**	0.900	0.897	0.894	0.891	0.955	0.953

and worse TPR compared with in-state applicants. These results indicate a strong structural preference towards in-state applicants.

Age, ethnicity, and continent of birth show moderate disparities. Younger applicants experience lower statistical parity (0.2358), and ethnic minority applicants exhibit both lower statistical parity and disparate impact. Geographic origin also plays a role, with applicants born in Africa, Asia, or Oceania experiencing substantially lower selection likelihood (disparate impact = 0.4659).

Overall, these results show that residency-based disparities are the most pronounced, while gender and military groups remain comparatively balanced. Age, ethnicity, and birthplace display intermediate levels of disparity that warrant attention but are not as severe as those of the Florida Residency group.

D. Deep-Dive into Florida Residency

A closer examination of the Florida Residency feature reveals why this attribute showed the strongest fairness disparities in the baseline model. The analysis compares selection rates and classification metrics across the privileged group (in-state applicants) and the unprivileged group (out-of-state applicants).

As shown in Fig 15, the original gap between selection rates of these groups was 42.8 percentage points. The model exacerbates the existing disparity even further since based on the predictions, the gap increases to 49.3 percentage points. This means that the model tends to admit an even higher proportion of in-state applicant and slightly under-predicts out-of-state admissions.

Group-level error analysis (Fig. 16) reveals clear systematic differences in how the model treats both groups. Although overall accuracy is similar, the underlying error patterns show meaningful bias.

The TPR is substantially higher for in-state applicants, meaning the model correctly admits them far more often. At the same time, the FPR for in-state students is also considerably higher, indicating that the model frequently over-predicts their admission, even when they do not qualify. In contrast, out-of-state applicants experience much higher FNR, meaning they are wrongly rejected more often even when they should be admitted. In our results, out-of-state applicants are roughly four times more likely to be incorrectly denied compared to in-state students. Their higher TNR further shows

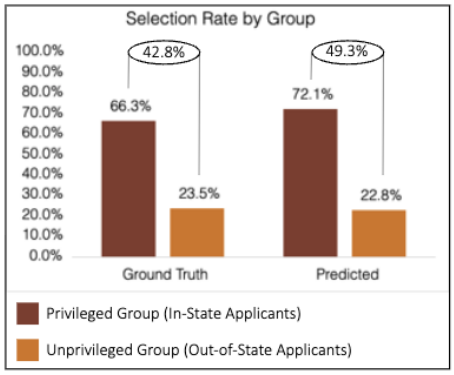

Fig. 15. Selection Rates per Florida Residency Groups (Ground Truth vs Predictions).

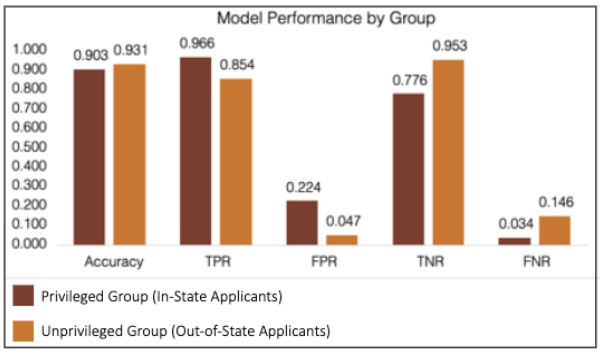

Fig. 16. Classification Rates per Florida Residency Groups

that the model is more conservative toward this group as it more readily assigning negative outcomes.

Taken together, these patterns indicate that the model systematically favors in-state applicants, both by granting them admissions they should not receive (high FPR) and by correctly admitting them at higher rates (high TPR). Conversely, out-of-state applicants face disproportionate barriers due to elevated false rejections, reinforcing the residency-based disparity observed in the fairness metrics.

E. Impact of Bias Mitigation

For the bias mitigation experiments, we focused on Florida Residency, as it was the attribute with consistently large disparities across all metrics.

TABLE IV
FAIRNESS EVALUATION METRICS ACROSS PROTECTED GROUPS

Category	Protected Group	Statistical Parity (↓)	Disparate Impact (↑)	Equal Opportunity (↓)	Equalized Odds (↓)
Gender	Female	0.0452	0.9156	0.0127	0.0443
Florida Residency	Out-of-State	0.4931	0.3158	0.1115	0.2877
Military Status	Non-Military	0.0538	0.9043	0.0832	0.1917
Age	Younger (15–28 yrs)	0.2358	0.6797	0.0446	0.2108
Ethnicity	Black/AA or Nonresident	0.2546	0.5761	0.0576	0.1369
Continent of Birth	Africa, Asia, or Oceania	0.2871	0.4659	0.0453	0.1500

Note: Best category (green) and worst category (red).

In the original training data, 78% of the admitted students were in-state compared to 22% out-of-state. Therefore, as outlined in the methodology, we explored two levels of pre-processing adjustments, a moderate re-balancing (70/30) and full parity (50/50) with both random oversampling and SMOTE. This section examines how these interventions affected both predictive performance and fairness outcomes.

1) Overall Model Performance: Across all four mitigation scenarios, the Random Forest model maintained strong predictive performance, with minimal changes compared to the baseline (Fig 17). Accuracy and F1-score varied by less than one percentage point across all re-balancing conditions, indicating that pre-processing interventions did not meaningfully degrade the model's ability to classify admitted vs. non-admitted applicants.

Moderate oversampling consistently preserved the strongest performance, producing accuracy (0.914–0.915), F1-score (0.914–0.915), and ROC-AUC (0.969) values nearly identical to the baseline. Full oversampling and SMOTE-based adjustments produced similar ROC-AUC values but resulted in slightly lower precision, suggesting that more aggressive re-balancing increases the risk of false positives.

Overall, these findings highlight that moderate re-balancing strikes the best balance between fairness gains and predictive reliability.

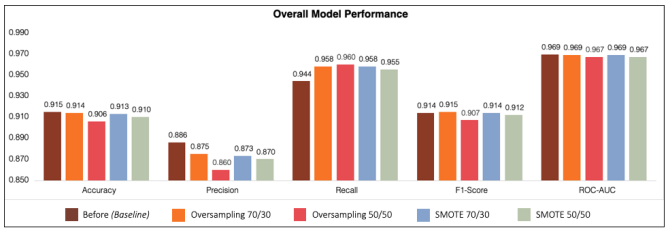

Fig. 17. Overall Model Performance Results Post Re-Balancing of Florida Residency

2) Fairness Metrics Comparison: Bias-mitigation interventions substantially improved fairness outcomes for the Florida Residency unprivileged group. Across both oversampling and SMOTE experiments, the interventions consistently narrowed group disparities. Statistical parity, equal opportunity, and equalized odds all moved closer to zero relative to the baseline, while Disparate Impact increased toward the desirable 0.8 threshold. Table V summarizes these results by showing the baseline metrics (e.g., before mitigation) and the percent change that resulted from each test.

Among all techniques, Random Oversampling (50/50) produced the largest improvements across fairness metrics, including a notable reduction in equal opportunity difference and increased disparate impact. Moderate re-balancing also meaningfully reduced disparities and improved multiple group-level error rates, while maintaining the strongest overall predictive performance. SMOTE offered improvements as well, but its gains were less consistent across metrics, reflecting limitations of synthetic sample generation in representing minority-group applicants.

Taken together, the fairness comparison shows that re-balancing the training data is effective in reducing residency-based disparities, with the 50/50 Random Oversampling condition achieving the most equitable outcomes, and the 70/30 split offering the best trade-off between performance and fairness.

VIII. CONCLUSION AND FUTURE WORK

This study evaluated the predictive performance and fairness of machine learning models applied to undergraduate admissions data, with a particular focus on identifying and mitigating disparities across demographic groups. All six models implemented demonstrated strong predictive performance. Fairness evaluations, however, revealed systematic disparities across several protected groups, most notably between in-state and out-of-state applicants. These patterns indicate that the observed disparities stem from underlying data characteristics rather than model-specific behavior. Importantly, demographic features were not among the strongest predictors of admission, which is an encouraging sign that the model is not directly relying on sensitive attributes.

To address disparity, we implemented a series of pre-processing bias mitigation strategies, including random oversampling and SMOTE at both moderate and balanced levels. Results showed that fairness can be significantly improved without harming overall predictive performance. These findings demonstrate that targeted pre-processing interventions can meaningfully increase equity in model outcomes without sacrificing model reliability.

Future work can be strengthened by expanding the dataset and exploring additional fairness-aware modeling strategies. As more information becomes available, incorporating additional demographic attributes (e.g., socioeconomic indicators,

TABLE V
FAIRNESS METRICS FOR FLORIDA RESIDENCY BEFORE AND AFTER MITIGATION

Metrics	Baseline		Oversampling 70/30		Oversampling 50/50		SMOTE 70/30		SMOTE 50/50	
FL Residency	Priv	Unpriv	Priv	Unpriv	Priv	Unpriv	Priv	Unpriv	Priv	Unpriv
Accuracy	0.903	0.931	0.00%	-0.17%	-0.10%	-2.06%	-0.10%	-0.29%	-0.10%	-1.09%
TPR	0.966	0.854	-0.23%	9.65%	-0.82%	13.55%	0.05%	8.42%	-1.00%	11.29%
FPR	0.224	0.047	-1.98%	54.26%	-5.95%	122.34%	1.59%	51.06%	-7.54%	86.17%
TNR	0.776	0.953	0.57%	-2.70%	1.71%	-6.09%	-0.46%	-2.54%	2.17%	-4.29%
FNR	0.034	0.146	6.41%	-56.63%	23.08%	-79.52%	-1.28%	-49.40%	28.21%	-66.27%
Statistical Parity (↓)	0.493		-8.38%		-16.34%		-6.77%		-13.24%	
Disparate Impact (↑)	0.316		17.34%		32.94%		15.08%		25.50%	
Equal Opportunity (↓)	0.111		-75.93%		-89.07%		-64.13%		-95.19%	
Equalized Odds (↓)	0.288		-39.91%		-59.30%		-32.03%		-56.95%	

Note: Improvement vs baseline (green), least worst vs baseline (orange).

parental education, or language status) would enable a more comprehensive assessment of equity across applicant subgroups. Increasing the number of admission cycles included in the analysis would also improve sample size and allow for evaluation of whether fairness patterns are stable over time or across cohorts. Finally, testing fairness-aware algorithms and post-processing methods such as equalized odds adjustments [7,9,15], would provide valuable comparisons against the pre-processing strategies used in this study and help identify the most effective techniques for reducing bias in admissions modeling.

Further evaluation should also include intersectional fairness analysis to examine combinations of attributes. For example, residency and ethnicity or birthplace, to uncover disparities that may remain hidden when each attribute is analyzed in isolation. Performance and fairness metrics could also be disaggregated by application type or intended college/major, as fairness dynamics often vary across academic units or program pathways. Additionally, assessing temporal trends across semesters or admission cycles would help determine whether fairness concerns change over time, providing valuable insight for institutional monitoring and long-term policy decisions.

LIMITATIONS

While this study provides meaningful insights into fairness and predictive modeling within undergraduate admissions, several limitations should be considered when interpreting the results.

Dataset and Institutional Context. The analysis is based on FIU's Fall 2024 applicant data, which reflects the unique demographic makeup and institutional policies of a large Hispanic-serving public university in Florida. As a result, patterns observed here may not generalize to other universities. For example, racial and ethnic distributions in Florida differ significantly from national averages. Groups commonly treated as underrepresented in many institutions (Hispanic/Latino applicants) represent a majority at FIU and therefore do not function as an unprivileged group in this context. Similarly, FIU's ethnicity categorization (non-resident alien) may obscure finer-grained subgroup differences.

External Policy and Societal Factors. Admissions outcomes in 2024 may have been influenced by geopolitical or policy-driven considerations outside the scope of this study. For instance, heightened scrutiny toward applicants from certain countries (e.g., China) could have affected selection rates, leading to discrepancies between "continent of birth" and "ethnicity" fairness patterns. These external influences could introduce bias that is dataset-specific and not representative of structural model behavior.

Feature Limitations and Data Quality Constraints. Certain features required assumptions due to missing or ambiguous information. High-school GPA values were recorded without consistent indication of scale (4.0 vs. 5.0 systems), limiting the precision of academic comparisons across applicants. Additionally, some demographic attributes may have been self-reported inconsistently, and missingness patterns could influence model behavior despite pre-processing efforts.

Policy-Driven Admission Constraints. FIU, like many public universities, operates under state-level policies that influence admissions decisions. For example, Florida's cap on out-of-state enrollment (often at 10%) may inherently disadvantage out-of-state applicants regardless of academic qualifications. Such policy constraints affect both ground-truth outcomes and predicted outcomes, making it difficult to disentangle model-driven bias from institutional constraints.

Generalization of Protected Group Definitions. Protected group definitions used in fairness evaluation (e.g., out-of-state vs. in-state, younger vs. older applicants) were tailored to this dataset. However, these definitions may differ in other states or institutions based on demographic composition, admissions priorities, or legal requirements. Consequently, fairness metrics might yield different interpretations if applied to universities with different protected-group structures.

State-Specific Legal and Policy Variability. Each U.S. state operates under its own regulations regarding admissions, demographic reporting, and fairness considerations. Florida's policies, both historical and current, shape admissions decisions in ways that may not apply elsewhere. Thus, replicating this study at another institution would require revisiting protected-group definitions, fairness criteria, and mitigation

strategies to ensure contextual relevance.

ACKNOWLEDGMENT

We would like to express our sincere gratitude to those who supported and guided this project. We thank Dr. Ananda Mondal, our course instructor, for providing direction, clear timelines, and ongoing feedback across both semesters of the capstone. We are especially grateful to Dr. Agoritsa Polyzou, who served as our project mentor and met with us regularly to review our progress, offer detailed technical and conceptual guidance, and help shape the development of this work from start to finish. We also extend our appreciation to Dr. Stephanie Lunn, our third evaluation committee member, whose expertise in educational systems and fairness provided valuable insights during the evaluation of our presentation and meaningful suggestions to strengthen the study. Their collective support greatly contributed to the successful completion of this project.

REFERENCES

[1] S. Barocas, M. Hardt, and A. Narayanan, *Fairness and Machine Learning: Limitations and Opportunities*. MIT Press, 2023. [Online]. Available: https://fairmlbook.org/

[2] N. Mehrabi, F. Morstatter, N. Saxena, K. Lerman, and A. Galstyan, "A survey on bias and fairness in machine learning," *arXiv preprint arXiv:1908.09635*, 2019. doi: 10.48550/arXiv.1908.09635.

[3] K. A. Bird, B. L. Castleman, and Y. Song, "Are algorithms biased in education? Exploring racial bias in predicting community college student success," *Journal of Policy Analysis and Management*, vol. 44, no. 2, pp. 379–402, 2024. doi: 10.1002/pam.22569.

[4] D. Gándara, H. Anahideh, M. P. Ison, and L. Picchiarini, "Inside the black box: Detecting and mitigating algorithmic bias across racialized groups in college student-success prediction," *AERA Open*, vol. 10, 2024. doi: 10.1177/23328584241258741.

[5] U.S. Department of Education, "U.S. Secretary of Education Linda McMahon directs National Center for Education Statistics to collect universities' data on race discrimination in admissions," Aug. 7, 2025. [Online]. Available: https://www.ed.gov/about/news/press-release/us-secretary-of-education-linda-mcmahon-directs-national-center-education-statistics-collect-universities-data-race-discrimination-admissions

[6] Lawyers' Committee for Civil Rights Under Law, "Summary of the Supreme Court's Decision in SFFA v. Harvard and SFFA v. UNC," Aug. 2023. [Online]. Available: https://www.lawyerscommittee.org/wp-content/uploads/2023/08/LC_Harvard-UNC-Cases_D.pdf

[7] M. Hardt, E. Price, and N. Srebro, "Equality of opportunity in supervised learning," *arXiv preprint arXiv:1610.02413*, 2016. [Online]. Available: https://arxiv.org/pdf/1610.02413

[8] G. Raftopoulos, G. Davrazos, and S. Kotsiantis, "Fair and transparent student admission prediction using machine learning models," *Algorithms*, vol. 17, no. 12, p. 572, 2024. doi: 10.3390/a17120572.

[9] K. Van Busum and S. Fang, "Interactive mitigation of biases in machine learning models for undergraduate student admissions," *AI*, vol. 6, no. 7, p. 152, 2025. doi: 10.3390/ai6070152.

[10] A. Castelnovo, R. Crupi, G. Greco, *et al.*, "A clarification of the nuances in the fairness metrics landscape," *Scientific Reports*, vol. 12, p. 4209, 2022. doi: 10.1038/s41598-022-07939-1.

[11] P. Domingos, "A few useful things to know about machine learning," *Communications of the ACM*, vol. 55, no. 10, pp. 78–87, 2012. doi: 10.1145/2347736.2347755.

[12] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed. Springer, 2009. [Online]. Available: https://hastie.su.domains/ElemStatLearn/

[13] Lumenova AI, "Fairness and bias in machine learning: Mitigation strategies," Jul. 23, 2024. [Online]. Available: https://www.lumenova.ai/blog/fairness-bias-machine-learning/

[14] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *Journal of Artificial Intelligence Research*, 2002. doi: 10.1613/jair.953.

[15] K. Varshney, "Introducing AI Fairness 360," IBM Research Blog, Sep. 19, 2018. [Online]. Available: https://research.ibm.com/blog/ai-fairness-360

[16] J. M. Irving, "Part 1: Creating dataset objects in AI Fairness 360 and exploring fairness metrics," Aug. 2, 2024. [Online]. Available: https://medium.com/@james.irving.phd/blog-post-series-ai-fairness-360-mitigating-bias-in-machine-learning-models-c1ec744c91c4

[17] J. M. Irving, "Part 2: Implementing bias mitigation techniques with AI Fairness 360," Aug. 2, 2024. [Online]. Available: https://medium.com/@james.irving.phd/blog-post-series-ai-fairness-360-mitigating-bias-in-machine-learning-models-2268e01584bd