A MACHINE LEARNING APPROACH TO REPLICATE INTERVIEW ASSESSMENTS FOR SELECTING AIR TRAFFIC CONTROLLER CANDIDATES

Author 1: *Mustafa Özdemir*[1]
Author 2. *Müjgan Sağır Özdemir*[2]

Highlights

- Machine learning models using a binary classification approach incorporate several candidate factors, such as exam results, high school background, and high school Grade Point Average.
- Machine learning models replicate ATCO interview outcomes, achieving up to 95% accuracy and enhancing objectivity in candidate selection.
- Results are compared with previous studies that use AHP/ANP-based approaches.

ABSTRACT

Selecting suitable air traffic control (ATCO) candidates is critical due to the demanding responsibilities of the role, which requires rapid decision-making and strong stress management. Traditional methods, such as multistage interviews, are commonly used in this selection process but face challenges, including subjectivity and lack of standardization. To address these issues, we explore machine learning (ML) as an alternative to streamline and improve the reliability of candidate evaluations. This study develops three ML models including Logistic Regression (LR), Support Vector Machine (SVM), and Decision Tree (DT) to replicate the outcomes of the interview phase. The candidate selection process was framed as a binary classification problem based on features such as previous exam results, high school background, and high school Grade Point Average. The results from the best-performing ML model are compared with previous studies that used AHP/ANP-based approaches. Our findings indicate that ML models can closely replicate interview exams, achieving accuracy rates of 95% for LR, 93% for DT, and 95% for SVM, highlighting their potential for practical application in ATCO candidate selection.

Keywords (3-6): ATCO candidate selection, Interview process, Machine learning.

[1] Mustafa Özdemir, Civil Aviation School, Erzincan Binali Yildirim University, Erzincan, Turkey e-mail: mozdemir@erzincan.edu.tr (ORCID: 0000-0002-0874-1044).

[2] Müjgan Sağır, Department of Industrial Engineering, Eskisehir Osmangazi University, Eskisehir, Turkey, e-mail: mujgan.sagir@gmail.com (ORCID: 0000-0003-2781-658X).

1. Introduction

Air Traffic Controller (ATCO) are responsible for maintaining the safe, orderly, and efficient movement of aircraft in airspace and at airports, where even small errors may lead to significant safety risks (Federal Aviation Administration [FAA], n.d.). Therefore, selecting qualified ATCO candidates is essential due to the critical responsibilities of the role, such as rapid decision-making and effective stress management. Traditionally, the ATCO candidate selection process relies on multistage evaluations, including interviews that assess skills like communication, stress control, and attention level. However, the interview process can be time-consuming and is subject to several limitations, such as subjectivity and a lack of standardization, which can hinder fairness and consistency in selection. These challenges make it difficult to ensure the selection of only the most suitable candidates, which may lead to higher training costs and even jeopardize safety standards.

In response to these limitations, this study proposes machine learning (ML) as a modern alternative to the traditional interview phase. Specifically, we investigate whether ML models can replicate interview assessments of ATCO candidates, using a binary classification approach that considers candidate-related features such as previous exam results, high school background, and Grade Point Average (GPA). Consequently, we frame our research question as follows: Can machine learning models offer a reliable and objective evaluation of ATCO candidates, comparable to traditional interview methods?

This study holds importance for multiple reasons. First, it aims to enhance ATCO selection accuracy and reliability, aligning with the aviation sector's high safety and performance standards. Second, it addresses operational inefficiencies by reducing reliance on time-intensive interview assessments, which may lower costs and expedite candidate selection. By developing and testing machine learning models capable of closely mirroring interview outcomes, this study provides a robust framework for ML-driven evaluations in ATCO selection and highlights potential for broader applications in high-stakes personnel selection across aviation and other safety-critical fields.

2. Literature Review

Limited studies have explored structured selection methods for ATCO candidates, despite the critical need for accuracy and objectivity in this profession. Analytic Hierarchy Process (AHP) and Analytic Network Process (ANP) have been applied in previous research to eliminate shortcomings in interview process for ATCO selection. Özdemir and Sağır Özdemir used AHP to support interview phase of ATCO selection process (Özdemir and Sağır Özdemir, 2018). In another study, they used ANP to determine the criteria weights and rank the candidates (Özdemir and Sağır Özdemir, 2021). In these studies, authors used several criteria such as communication skills, stress control, and attention level. Taylan et al. framed ATCO selection as a multi-criteria decision-making problem, aiming to choose the best candidate among ten alternatives based on five criteria: communication, math skills, awareness, personal skills, and motivation. They used a fuzzy decision tree to weight these criteria based on five decision-makers' evaluations, followed by fuzzy TOPSIS to rank the candidates (Taylan, 2014). However, these studies do not fully eliminate the interview phase; instead, they use these techniques as supplementary tools, still requiring the interview to be conducted. This study is one of the first to explore ML as an alternative approach for automating and enhancing consistency in ATCO candidate assessments. It addresses this gap by employing ML models to replicate interview evaluations, offering a scalable, objective solution that reduces human subjectivity.

3. Objectives

The primary objective of this study is to develop a machine learning-based approach to replicate the interview assessments in the selection process for ATCO candidates, aiming to improve objectivity, consistency, and scalability.

The hypotheses are that (1) machine learning models can closely replicate interview outcomes with high accuracy and (2) they provide a more objective and standardized approach, potentially outperforming traditional methods in reliability.

4. Methodology

The models chosen for this study include Logistic Regression (LR), Support Vector Machine (SVM), and Decision Tree (DT), each selected for its specific strengths in classification tasks. LR is well-suited for binary classification problems and provides interpretable outputs, allowing us to analyze the relationship between candidate attributes and suitability. SVM, known for its effectiveness in handling complex, non-linear relationships, was chosen for its robustness and accuracy in separating candidate categories. Finally, DT was selected for its interpretability and transparency, enabling us to visualize the decision-making process in a structured way, which can be valuable for understanding the reasoning behind each candidate's classification.

The dataset consists of 194 candidates, with only 30 passing the interview phase. Accordingly, we encoded our target variable, assigning 1 to successful candidates and 0 to others. Since only 30 candidates passed the interview exam, this results in an imbalanced distribution between the two classes (i.e., selected and eliminated), with a ratio of approximately one to five. To address this issue, we applied oversampling techniques. Oversampling involves increasing the number of instances in the minority class by duplicating existing examples or generating synthetic data. This approach helps to balance the class distribution, ensuring that machine learning models can learn more effectively and avoid biases towards the majority class.

The classification models use a total of 16 selected features, including gender, previous exam results (covering six types of grades that assess various aspects of candidates' backgrounds, such as science, social science, literature, and their combinations), high school domain, high school GPA, and city of birth. Some features, such as city of birth, were selected for exploratory purposes, while others were chosen for their relevance to core competencies and academic performance, which can indicate a candidate’s potential for success in ATCO training programs. The city of birth is coded as a binary variable, set to 1 if the candidate is from the same city as the school. This feature may hold significance, as most candidates come from the school’s city. The school’s limited recognition in other cities and the travel distance may discourage applicants from outside areas. Using these features, the models aim to objectively assess candidates based on their history and performance metrics.

To evaluate the performance of each model, we used several key metrics, including accuracy, precision, recall, and F1-score. Accuracy measures the overall correctness of the model's predictions, while precision and recall provide insights into the model's ability to correctly identify positive cases (suitable candidates). The F1-score, a harmonic mean of precision and recall, was also calculated to provide a balanced assessment of model performance, particularly given the imbalanced nature of the dataset.

5. Results

The three machine learning models were evaluated for their effectiveness in replicating ATCO interview assessments. LR achieved an overall accuracy of 95%, with strong performance across both classes (weighted F1-score of 0.95), demonstrating its ability to distinguish between suitable and unsuitable candidates. The DT model had a slightly lower accuracy of 93% and a weighted F1-score of 0.93, with reduced precision for the positive class (0.77), indicating some inconsistency in classifying suitable candidates. SVM produced results similar to LR, with a 95% accuracy and a weighted F1-score of 0.95.

Overall, LR and SVM showed high accuracy and balanced performance, making them suitable alternatives for ATCO candidate assessments. The DT model showed slightly lower precision in identifying suitable candidates.

Table 1. Performance metrics for machine learning models

Model	Accuracy	Precision (Class 1)	Recall (Class 1)	Weighted Avg F1-Score
LR	0.95	0.83	0.91	0.95
DT	0.93	0.77	0.91	0.93
SVM	0.95	0.83	0.91	0.95

Further analysis was conducted on the LR model to assess its ability to correctly classify selected and eliminated candidates, as it is one of the best-performing models. The performance of the LR was evaluated using a confusion matrix, as shown in Figure 1. The matrix provides a detailed breakdown of the model's predictions across the two classes: selected and eliminated candidates.

- The model correctly identified 46 eliminated candidates as eliminated (true negatives) and correctly classified 10 selected candidates as selected (true positives).

- However, the model incorrectly predicted 2 eliminated candidates as selected (false positives) and 1 selected candidate as eliminated (false negatives).

These results indicate that the model is highly accurate in distinguishing eliminated candidates but is slightly less consistent in identifying selected candidates. This imbalance reflects the inherent challenge of working with an imbalanced dataset, where the selected class constitutes a smaller portion of the data.

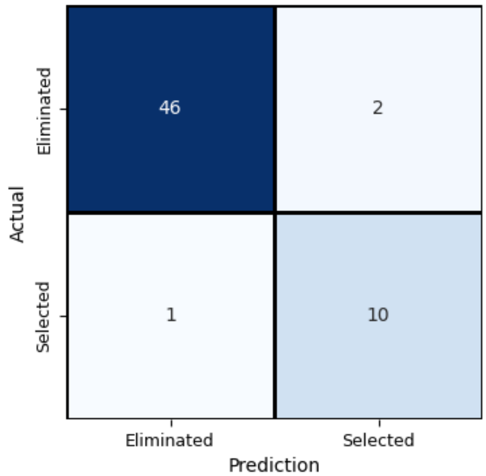
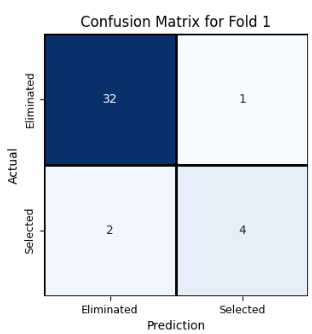
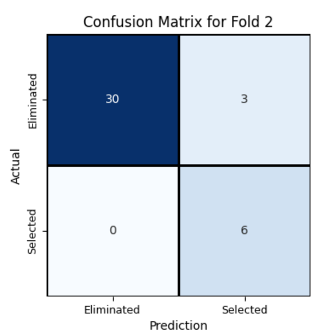


Figure 1. Confusion matrix for logistic regression model showing actual vs. predicted classifications.

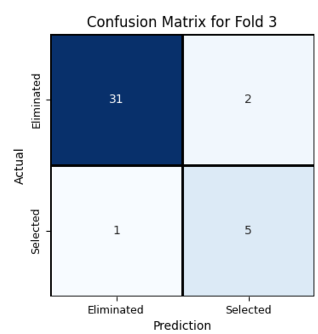
We explored the model's generalizability across different sub-datasets using 5-fold cross-validation, with the following performance results. For precision (Class 1), the average is 0.6929 ± 0.0380, indicating moderate precision in correctly identifying Class 1. The recall (Class 1) is high, with an average of 0.9000 ± 0.1491, although there is some variability across the folds. The weighted average F1-score is 0.7786 ± 0.0702, suggesting a good balance between precision and recall, but still room for improvement. The model achieves a good accuracy of 0.9227 ± 0.0182, reflecting strong overall performance. These results show that while the model performs consistently well, there is some variation in its ability to precisely classify Class 1 instances. Figure 2 presents confusion matrix for each fold of 5-fold cross-validation. As observed, while the model performs reasonably well overall, fold 4 (Figure 2-d) fails to classify five of the eliminated candidates correctly.



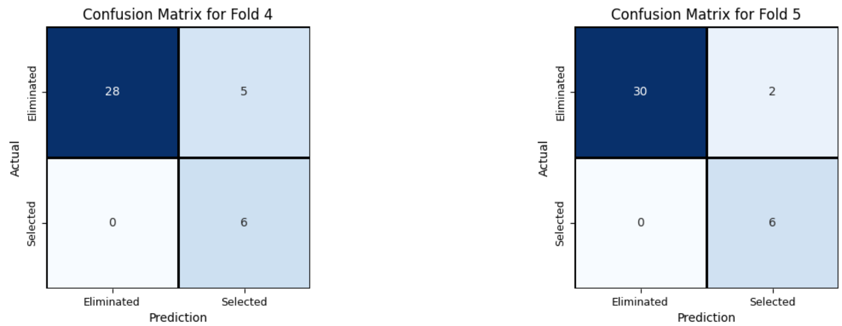
(a)



(b)



(c)



(d) (e)

Figure 2. Confusion matrices for each fold of 5-fold cross-validation using logistic regression

Figure 3 illustrates the performance of LR in identifying selected and eliminated candidates. Only one candidate was incorrectly categorized as eliminated, despite being selected in the actual interview. Additionally, two candidates were classified as eliminated but were actually selected. This resulted in a total of three misclassifications. Comparing these results to those of Özdemir (2021), who applied an ANP approach to the same dataset, their model misclassified 4 out of 30 candidates, resulting in an accuracy of 0.87. Based on the test data, only 1 out of 11 candidates was misclassified, resulting in a higher accuracy of 0.91. While there may be slight variations in misclassification for other candidates not shown in the figure, it is important to note that these results were achieved without the need for an interview, demonstrating the potential effectiveness of this approach.

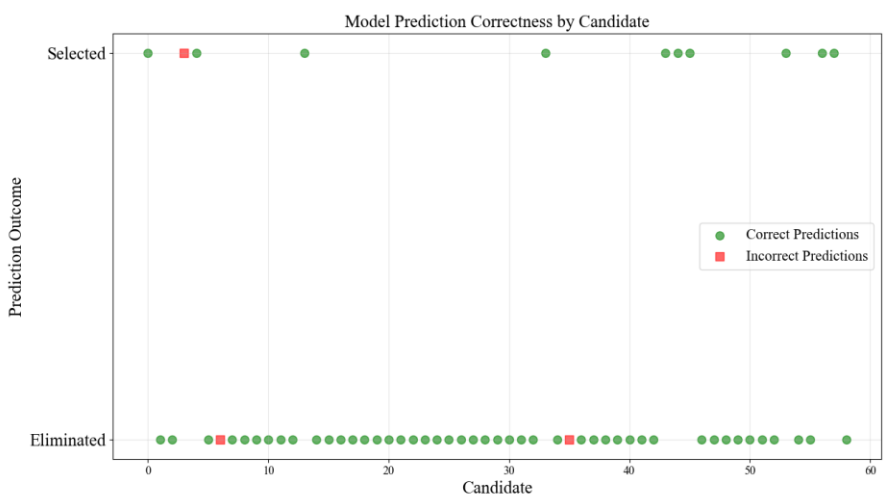


Figure 3. LR performance on test data for correctly identifying selected and eliminated candidates

The feature importance results in Figures 4 and 5 highlight the key attributes for predicting selected candidates in both SVM and LR models. For SVM, the three most influential features are aptitude test, city of birth, and gender. This suggests that candidates' performance on the aptitude test is a strong differentiator, with city of birth and gender also playing significant roles. In LR, the top three features are city of birth, high school domain, and aptitude test. Again, city of birth emerges as a significant factor, aligning with the SVM results. The emphasis on city of birth indicates that location may be a factor in candidate selection, possibly due to local familiarity with the school or accessibility. While this may initially seem counterintuitive, closer examination of the data reveals that 19.6% of candidates from the university's city were selected, compared to just 9.8% of candidates from outside the university's city. This disparity highlights the potential role that geographic factors may play in the selection process. The high school domain, which may relate to candidates' educational backgrounds or areas of focus, and aptitude test scores further provide a strong basis for assessing suitability.

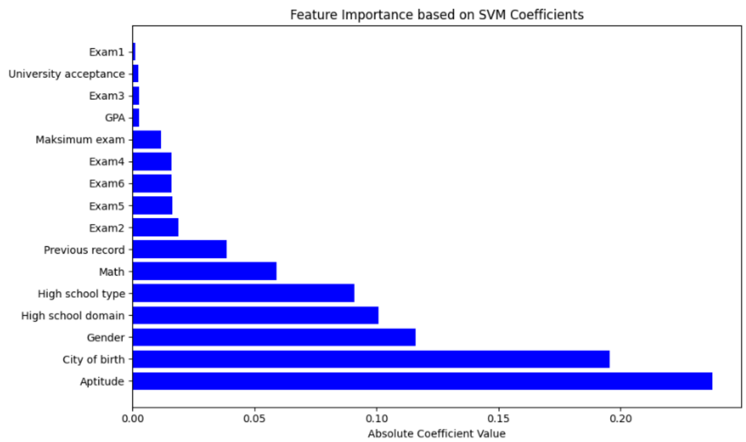


Figure 4. Feature importance based on SVM

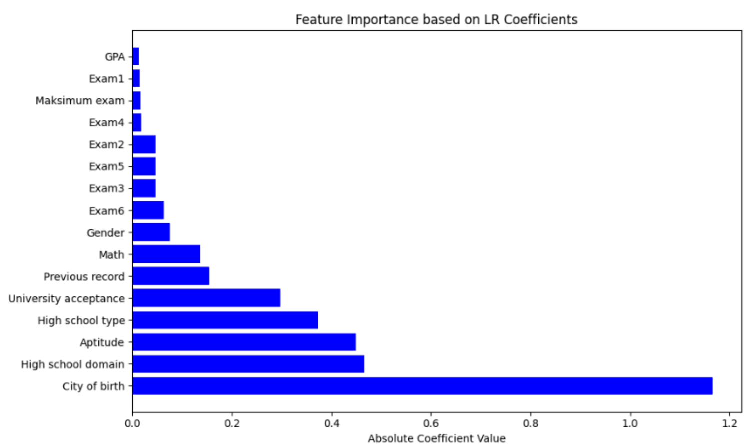


Figure 5. Feature importance based on LR

6. Conclusions

In conclusion, this study demonstrates the potential of machine learning as a viable alternative for enhancing the ATCO candidate selection process. By integrating SVM and LR models, the approach offers a scalable, objective, and consistent method that replicates key aspects of the traditional interview phase while minimizing human subjectivity. Despite minor misclassifications, the models achieved promising accuracy in identifying both selected and eliminated candidates, suggesting that ML can effectively support ATCO selection by streamlining candidate assessment and reducing reliance on time-intensive interviews. This study lays the groundwork for further development and refinement, ultimately contributing to a more efficient and reliable ATCO selection framework.

7. Limitations

While this study demonstrates the potential of machine learning models to assess ATCO candidates, several limitations must be acknowledged. First, the data size is limited, with data available for only 194 candidates. A larger dataset would improve model robustness and provide a broader basis for generalization, potentially enhancing the reliability of the results. Second, since this study used interview grades as the target variable, there is a risk that the machine learning models may have inadvertently learned and reproduced any subjectivity or biases present in the original interview assessments. While our goal was to enhance objectivity, the models may still reflect implicit biases within the historical data, potentially impacting fairness in candidate selection. This limitation could be mitigated by incorporating additional candidate characteristics not available in the current dataset. Adding features such as psychometric assessments or cognitive test scores could help the models capture the nuanced competencies essential for ATCOs, thereby increasing objectivity and fairness. Despite these limitations, this study contributes valuable insights into the use of machine learning for ATCO selection, providing a foundation for future research and potential refinements in candidate evaluation processes.

8. Key References

Federal Aviation Administration (FAA). (n.d.). *Section 5. Pilot/controller roles and responsibilities*. FAA. Retrieved November 23, 2024, from https://www.faa.gov/air_traffic/publications/atpubs/aim_html/chap5_section_5.html.

Taylan, O., Alidrisi, H., & Kabli, M. (2014). A multi-criteria decision-making approach that combines fuzzy topsis and DEA methodologies. *South African Journal of Industrial Engineering*, 25(3), 39-56.

Özdemir, M., & Sağır Özdemir, M. (2021). Application of the ANP in the interview phase of air traffic controller candidate selection process. *Aircraft Engineering and Aerospace Technology*, 93(10), 1616-1623.

Özdemir, M., & Sağır Özdemir, M. (2018). The use of the analytic hierarchy process for admission to an air traffic control department. *International Journal of the Analytic Hierarchy Process*, 10(1).