



Article

Using Machine Learning with Eye-Tracking Data to Predict if a Recruiter Will Approve a Resume

Angel Pina¹, Corbin Petersheim¹ , Josh Cherian¹ , Joanna Nicole Lahey^{2,*} , Gerianne Alexander³ and Tracy Hammond¹

¹ Sketch Recognition Lab., Department of Computer Science & Engineering, Texas A&M University, College Station, TX 77843, USA; jcherian14@tamu.edu (J.C.); hammond@exchange.tamu.edu (T.H.)

² Bush School of Government and Public Service, Texas A&M University, College Station, TX 77843, USA

³ Department of Psychology, Texas A&M University, College Station, TX 77843, USA; galexander@tamu.edu

* Correspondence: jlahey@tamu.edu; Tel.: +1-(979)-458-3463

Abstract: When job seekers are unsuccessful in getting a position, they often do not get feedback to inform them on how to develop a better application in the future. Therefore, there is a critical need to understand what qualifications recruiters value in order to help applicants. To address this need, we utilized eye-trackers to measure and record visual data of recruiters screening resumes to gain insight into which Areas of Interest (AOIs) influenced recruiters' decisions the most. Using just this eye-tracking data, we trained a machine learning classifier to predict whether or not a recruiter would move a resume on to the next level of the hiring process with an AUC of 0.767. We found that features associated with recruiters looking outside the content of a resume were most predictive of their decision as well as total time viewing the resume and time spent on the Experience and Education sections. We hypothesize that this behavior is indicative of the recruiter reflecting on the content of the resume. These initial results show that applicants should focus on designing clear and concise resumes that are easy for recruiters to absorb and think about, with additional attention given to the Experience and Education sections.

Keywords: machine learning; resumes; eye-tracking; recruiter



Citation: Pina, A.; Petersheim, C.; Cherian, J.; Lahey, J.N.; Alexander, G.; Hammond, T. Using Machine Learning with Eye-Tracking Data to Predict if a Recruiter Will Approve a Resume. *Mach. Learn. Knowl. Extr.* **2023**, *5*, 713–724. <https://doi.org/10.3390/make5030038>

Academic Editor: Andreas Holzinger

Received: 17 May 2023

Revised: 7 June 2023

Accepted: 22 June 2023

Published: 28 June 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

A resume serves as a tool for students to showcase their candidacy to recruiters [1]. Candidates must convey that they possess the requisite skills and abilities to excel in the position to which they are applying. Often, the first opportunity for an applicant to impress a potential employer is the resume [2]. A common practice used by employers is to use recruiting software to filter applicants prior to being reviewed by a recruiter [3]. While helpful, 60% of recruiters believe that qualified candidates are filtered out by the software [4]. Even when a resume makes it beyond this stage, resume screening by a recruiter is a brief process, lasting on average between thirty seconds and three minutes [5]. Therefore, it is important for students to be able to quickly and effectively convey their skills through their resume to be competitive in the hiring process. To accomplish this, students must understand which areas on the resume contribute most to their application in the eyes of a recruiter.

Science, technology, engineering, and mathematics (STEM) jobs are vital to a nation as they impact economic growth and standard of living [6]. As such, we must assist qualified STEM graduates in finding jobs for the betterment of the field. Within the STEM field, positions for computer science graduates are competitive [7], creating a barrier to entry for recent graduates attempting to enter the workforce. Considerable emphasis is placed on previous experience when applying for computer science positions [8], but it is difficult to obtain this experience without a well-constructed resume. Though there is a sizeable body of research in resume construction [9], this research is primarily focused on majors

outside of STEM and may not be applicable to computer science graduates. Additionally, much of this past resume research relies on surveys [9] rather than actually simulating the resume screening process. By better understanding this screening process and focusing on computer science resumes, we seek to aid recent computer science graduates in creating effective resumes.

Eye-tracking and machine learning provide new avenues from which to explore the resume screening process. Eye-tracking has been widely used across a number of fields to provide greater insight into what individuals are paying attention to on a system or interface. Here, we leverage it to understand what parts of the resume are of the greatest importance to their decision-making process. Machine learning has become even more ubiquitous as researchers and practitioners have demonstrated that these techniques can be used to accurately model and predict behaviors, trends, and outcomes across nearly every field of interest. Our study utilizes these tools to make two contributions. First, we present a machine learning pipeline that predicts whether or not a resume will be moved to the next level of the hiring process based on the eye movements of the recruiter. Second, by determining which sections of computer science resumes are most predictive of a resume being moved to the next level, we provide insights into which sections contribute the most to entry-level computer science graduates in the eyes of recruiters.

2. Prior Work

2.1. Evaluating Resumes

A typical computer science resume consists of several vital sections, such as education, work experience, and technical skills [10]. In addition to these sections, projects and extracurriculars may also warrant inclusion on computer science resumes. While the results of past resume studies are mixed, they provide insight into what recruiters might be looking for during the resume screening process [9]. Below, we briefly address the importance of various resume sections based on the results of past studies in computer science and other fields.

2.1.1. Academic Qualifications

The academic qualifications section on a resume can be taken as a combination of a candidate's education, GPA, and relevant courses. Having relevant educational credentials to the position being applied is of particular importance because even a well-formatted resume cannot compensate for a weak or irrelevant education [11]. A relevant education produces more positive perceptions of applicants [11] and influences the recruiters' perception of other academic qualifications such as GPA [12].

By far the most thoroughly researched academic qualification is GPA. GPA has been documented to influence recruiters' perceptions of applicant fit [13,14], cognitive ability [12,15,16], employability [17], motivation [15], and work ethic [13]. Although a higher GPA is generally preferred, including GPA even if it is lower is preferable to excluding GPA for entry-level positions [2]. Oftentimes, recruiters utilize GPA as a pre-selection criterion wherein a minimum GPA is set. Applicants falling below this threshold are dismissed from further consideration regardless of other qualifications [15,16,18]. Less clear is how well GPA is able to predict future job performance [19]. Regardless, GPA is typically considered when making recruiting decisions for positions in IT areas [18].

Relevant coursework is perhaps the least well-researched of the components of academic qualifications. The literature has mixed recommendations in regard to its inclusion, ranging from unimportant [20] to improving the likelihood of obtaining an invitation to interview [2].

2.1.2. Work Experience

The attribute on resumes that is most closely associated with future job performance is work experience [21]. Hiring managers often place greater weight on work experience than other credentials such as academic qualifications [22]. Accordingly, work experience

has consistently been regarded as a vital component of resumes [5,20,23–25]. However, not all work experience is created equal. Job experience relevant to the position being applied is of particular interest to employers when determining the employability of a candidate [11,13]. Even still, irrelevant work experience may contribute positively to the quality of an application [1,25]. For candidates applying for entry-level positions, this experience oftentimes comes in the form of internships. In a field experiment across various majors applying for entry-level positions, applicants with internship experience were 12.6% more likely to be invited to interview [26]. In computer science literature as well, previous experience, regardless of whether in the form of internship or industry position, was consistently rated highly in terms of importance for applicant quality [27,28].

2.1.3. Extracurriculars

Resume literature focuses considerably less on the effects of extracurriculars on resumes. Extracurricular activities as discussed here include both clubs as well as groups such as fraternities or sororities. What is typically agreed upon, though, is that they factor in positively to the evaluation of resumes [5,24]. Particularly impressive to employers is the presence of leadership positions within these activities [29,30]. Additionally, the number of activities participated and whether or not the activity was applicable to an applicant's career both contribute to applicant quality [30]. Recruiters may view extracurricular activities positively due to inferences drawn from their inclusion. Studies suggest that recruiters associate extracurriculars with interpersonal skills in candidates [12,31].

2.2. Eye-Tracking

Eye-tracking is a technique wherein both the movement and sequence of a subject's eyes are measured and recorded to provide objective data [32]. In virtually every scientific domain, eye-trackers provide a means to investigate underlying visual processes through the collection of quantitative information [33]. This technique can capture even subtle human behaviors. For instance, from eye-tracking data alone, researchers have been able to distinguish between participants recalling and imagining an event [34]. There have been few instances of researchers applying eye-tracking to resumes. One notable study applies eye-tracking research regarding computer screen reading patterns to derive a set of best practices for resume construction [35]. Another noteworthy study uses eye-tracking and resumes to detect recruiter discrimination based on the age, race, and gender apparent from the resume [36]. The present experiment differs from the aforementioned studies in that we are attempting to gauge whether the way in which a recruiter screens a resume is indicative of their decision regarding that resume.

2.3. Machine Learning

Machine learning is a subfield of artificial intelligence focused on the detection of patterns in data [37]. Though machine learning has an impressive array of possible applications that range from computer vision to natural language processing, little research has applied machine learning techniques to resumes. The research that does exist is dedicated to using machine learning in applicant filtering systems [38–40]. While likely helpful for companies with large numbers of applicants, this does little to aid students in how they should construct their resumes. The aim of this paper is to utilize machine learning to help students identify which sections on the resume factor the most into hiring decisions.

3. Research Methods

3.1. Study Recruitment

We collected data from 221 recruiters across various industries that hired computer science majors. Participants were found through STEM career fairs and businesses in a southwestern state. The study was conducted in private booths with a computer and a Tobii Spectrum eye-tracker. The eye-tracker was non-invasive and was attached to the computer monitor. Prior to the experiment, participants were given consent forms that

explained the task, their right to stop the study at any time with no consequences, and their ability to ask questions before proceeding. The study was completed on the computer using a mouse and keyboard. No time limit was given, and resumes could be reviewed as long as necessary. All participants were paid \$50 for their cooperation.

3.2. Experiment Process

Tobii Spectrum eye-trackers were calibrated to participants' eyes using iMotions and Inquisit by having participants follow dots on the screen with their eyes. First, iMotions was calibrated with a nine-point display, followed by Inquisit with a five-point display. After calibrating, participants were instructed on the study process and shown instructions on the screen.

Participants were first shown five practice resumes to familiarize themselves with the process. For each resume, participants had the option to check a box to determine whether or not they would move the resume to the next level of the hiring process. The first phase involved 30 resumes, each shown one at a time. After this, recruiters had the opportunity to cycle back through the 30 resumes if they wanted to do so. Following the completion of this phase, participants were asked what position they were thinking of while assessing the resumes.

For the second round, recruiters were once again shown the same 30 resumes, only this time they answered three questions: "rate the quality or 'hireability' of the previous candidate," "what type of position do you think this candidate will most likely end up?" and "what starting salary would you guess that this candidate would receive?". After the eye-tracking portion of the study was completed, participants filled out a short demographics survey on an iPad. The data collected have no identifying information, keeping participants anonymous.

To prepare our data for analysis, we organized and cleaned the data from the 221 participants. Only the data from the first round of resumes was used for this study because this round simulates the resume screening process. We removed the data of 24 participants that were incomplete due to hardware or software malfunction. To keep our findings consistent, we worked only with data from resumes that were complete. For a resume to be considered complete, it needed to have at least one piece of information on each of the seven sections included on all resumes in the study. We were able to use data from 2043 resumes following this step. Of these resumes, recruiters passed along 1257 (61.5%) to the next stage of the hiring process and did not pass along 786 (38.5%). For our analysis, we treated the resumes that were passed along as positive examples, and the resumes that were not passed along as negative examples.

3.3. Data Labeling

Resumes were divided into eight sections, as shown in Figure 1. The AOI from top to bottom were as follows: Introduction, Address, Education, Experience, Projects, Membership, Skills. The last area of interest is Outside, containing the space where there are no other sections or information to be seen.

Our study used 115 features to train our classifier as each of them dealt with different aspects of a recruiter's gaze. For each AOI, the following features were collected: gaze points, number of fixations, number of dwells, dwell duration, dwell rate, dwell duration average, and 'AOI from AOI'. The feature 'AOI from AOI' detects when there is a transition from one AOI to another AOI, including itself, resulting in 64 pairs. These features capture the saccades in the scanpath. The remaining three features dealt with the recruiter's gaze from the start to the end of their review of the resumes: fractal dimension, fractal dimension average, and stimulus duration. The fractal dimension represents the complexity of eye movement [41]. To find the value, we use the formula as defined in Equation (1), where ϵ is the length of the boxes, G is the gaze scanpath, and $N(\epsilon)$ is the number of boxes with

length ϵ to cover G . In our study, the fractal dimension feature analyzes the complexity of the recruiter’s eye movement throughout the resume (D).

$$D_{box}(G) := \lim_{\epsilon \rightarrow 0} \frac{\log N(\epsilon)}{\log(1/\epsilon)} \tag{1}$$

The full list of features and their definitions are listed in Table 1. We chose these features specifically to understand where the recruiter was looking and how that impacted their decision. To that end, we placed an emphasis on features that captured where and how long recruiters were looking. We did not utilize features that utilized pupil diameter because the lighting in our experiment was not controlled enough to make that data viable.

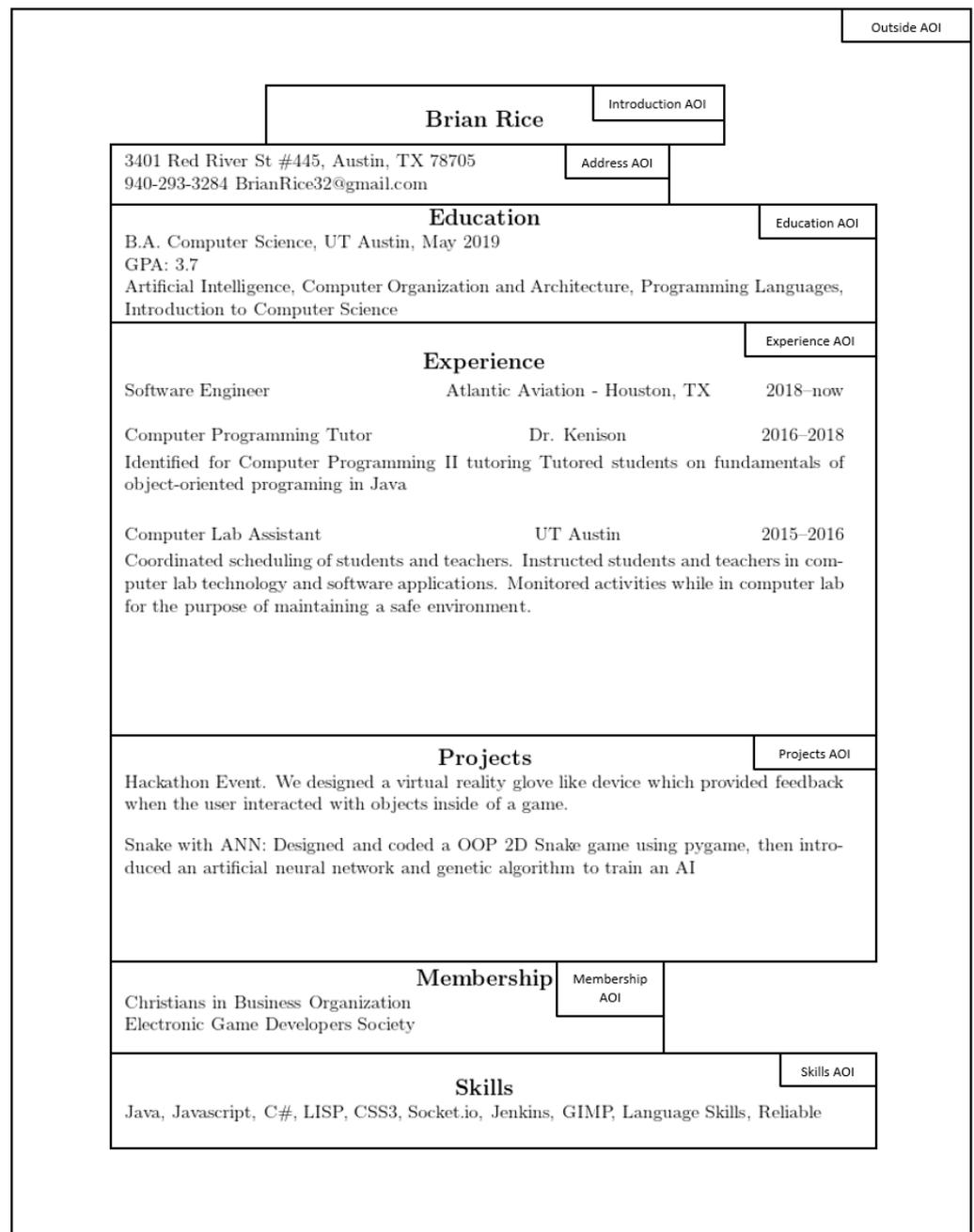


Figure 1. Resume sections with AOIs boxed and labeled.

Table 1. Definitions of Features Collected.

Feature Name	Definition
Gaze Points: X ¹	Samples taken by the eye tracker in screen coordinates
Number of Fixations: X ¹	Count of when gaze points are near and around each other for 100–300 milliseconds
Number of Dwells: X ¹	Count of when there are multiple fixations on one AOI ending with a fixation on another AOI
Dwell Duration: X ¹	Total time duration spent on dwells
Dwell Rate: X ¹	Number of dwells per the time spent looking at the AOI
Dwell Duration Average: X ¹	Average time spent on the AOI per dwell
X ¹ From Y ²	Count of transitions from one AOI to another AOI
Fractal Dimension	Complexity of the eye movements in a resume
Stimulus Duration	Total time spent looking at the resume

¹ X represents one of the eight AOIs. ² Y represents one of the eight AOIs, including X.

3.4. Machine Learning

We trained and tested various machine learning classifiers to identify which variables were most highly correlated with our outcome of interest, i.e., moving on to the next level of the hiring process. We used the scikit-learn [42] implementations of several common algorithms, including the Majority Classifier (called the Dummy Classifier in scikit-learn), Naive Bayes, K-Nearest Neighbors, Decision Tree, Random Forest, AdaBoost, Gradient Boosting, Support Vector Machine (SVM), and Multilayer Perceptron. These algorithms were selected to cover a variety of classifier types as they use distinct learning strategies. Each classifier used the default parameters, e.g., k being 5 for the K-Nearest Neighbors classifier. The Majority Classifier serves as a baseline since it always chooses the most common class with no input from the data. Naive Bayes determines the most likely class using maximum a posteriori estimation and makes the naive assumption that all the features are conditionally independent. K-Nearest Neighbors finds the k samples with minimum distance to the input in the feature space and selects the class via vote. Decision Tree generates a hierarchy of rules with maximum information gain based on the feature values to sort samples into a class. Random Forest is an ensemble method that aggregates the outcomes of many Decision Trees that were each trained on different bootstrapped samples of the dataset. AdaBoost is a boosting algorithm that fits a sequence of weak learners and determines the class based on a weighted vote. Gradient Boosting is also a boosting algorithm like AdaBoost, and it works with differentiable loss functions. SVM learns a number of decision boundaries (called support vectors) that separate the data into classes with maximal margin. Multilayer Perceptron is a neural network that learns a nonlinear function to map the feature values to the most likely class.

To evaluate the performance of our classifiers, we use the standard metrics of accuracy, precision, recall, F1-score, and the Area Under the Curve (AUC) of the Receiver Operating Characteristic (ROC). Each of these metrics is defined in terms of true positives (TP), false positives (FP), false negatives (FN), and true negatives (TN), where ‘true’ predictions are correctly labeled as the positive class or the negative class and ‘false’ predictions are errors. Accuracy gives the rate of correct predictions (Equation (2)). To gain further insights into the types of errors the classifier may incur, precision, recall, and F1-score are often reported with or in place of accuracy. Precision gives the positive predictive value, which is the opposite of the false discovery rate and is impacted by the presence of FP (Equation (3)). Recall is the true positive rate, which is also called sensitivity or hit rate, and is impacted by the presence of FN (Equation (4)). F1-score is the harmonic mean of precision and recall (Equation (5)).

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN} \quad (2)$$

$$\text{Precision} = \frac{TP}{TP + FP} = 1 - \frac{FP}{FP + TP} \quad (3)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (4)$$

$$\text{F1-Score} = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} = \frac{2TP}{2TP + FP + FN} \quad (5)$$

Each of the aforementioned metrics assumes a singular classification threshold. To report performance in a threshold-independent fashion, ROC plots the true positive rate, also known as recall, against the false positive rate, the opposite of the true negative rate (Equation (6)), over all possible thresholds. This graph is summarized into a metric by calculating the AUC; this metric is bounded between 0 and 1, where 1 means the classifier is always correct.

$$\text{False Positive Rate} = \frac{FP}{FP + TN} = 1 - \frac{TN}{TN + FP} \quad (6)$$

4. Results

In an effort to understand what parts of the resume were most important to recruiters, we first conducted our analysis without any of the features extracted from areas outside any of the resume sections. Specifically, we ran Leave-One-Subject-Out (LOSO) cross-validation for each of the classifiers we evaluated; the results can be seen in Table 2. The results shown are the average performance across cross-validation folds, followed by the standard deviation. Classifiers are listed in order of descending performance in terms of AUC, our metric of success. Of the classifiers tested, Random Forest performed the best; however, the performance was not good enough to draw reasonable conclusions.

Table 2. Average Performance Metrics Across Cross-Validation Folds Without Outside Features. Results are ordered by average AUC.

Classifier	Accuracy	F1	Precision	Recall	AUC
Random Forest	0.629 ± 0.182	0.692 ± 0.177	0.667 ± 0.217	0.595 ± 0.181	0.595 ± 0.181
Gradient Boosting	0.613 ± 0.161	0.681 ± 0.177	0.649 ± 0.214	0.571 ± 0.175	0.571 ± 0.175
AdaBoost	0.605 ± 0.154	0.667 ± 0.181	0.644 ± 0.212	0.562 ± 0.159	0.562 ± 0.159
Decision Tree	0.567 ± 0.168	0.610 ± 0.191	0.638 ± 0.236	0.560 ± 0.178	0.560 ± 0.178
Naive Bayes	0.542 ± 0.180	0.511 ± 0.241	0.655 ± 0.291	0.554 ± 0.176	0.554 ± 0.176
K-Nearest Neighbors	0.568 ± 0.164	0.633 ± 0.187	0.611 ± 0.223	0.529 ± 0.166	0.529 ± 0.166
Multilayer Perceptron	0.541 ± 0.189	0.553 ± 0.255	0.599 ± 0.288	0.525 ± 0.182	0.525 ± 0.182
SVM	0.617 ± 0.232	0.734 ± 0.203	0.617 ± 0.232	0.500 ± 0.000	0.500 ± 0.000
Majority	0.617 ± 0.232	0.734 ± 0.203	0.617 ± 0.232	0.500 ± 0.000	0.500 ± 0.000

Subsequently, we ran LOSO cross-validation for each of the classifiers with all of the features (i.e., including the features calculated for areas outside of specific resume AOIs). The results of this analysis are shown in Table 3. As with the previous results, Random Forest was the best-performing classifier. As mentioned, this model uses the default parameters in scikit-learn, which include 100 estimators, gini criterion, and no maximum depth, and only a subset of the features are considered at each split (specifically, \sqrt{m} , where m is the number of features).

To obtain further insight into the performance of our model, we conducted some additional analysis. Using Carrington et al.'s Deep ROC metric, we analyzed the model's performance at different thresholds of prediction risk to demonstrate how the model performs at varying classification thresholds. For this analysis, we looked at the performance of our model over four different groups: full range, high risk, medium risk, and low risk. These results are shown in Table 4. When few false positives are allowed, the model rejects approximately half of the candidates that would have moved onto the next stage, with an average recall or sensitivity of 0.651. As expected, performance is more balanced in the

medium category with an AUC of 0.814, precision of 0.688, and recall of 0.855. However, specificity decreases to 0.245, meaning about 1.5 times the candidates are moving onto the next stage compared to what recruiters would have actually selected. When many false positives are allowed, recall increases to 0.94, while specificity drops to 0.016, so the model rejects very few candidates.

Table 3. Average Performance Metrics Across Cross-Validation Folds. Results are ordered by by average AUC.

Classifier	Accuracy	F1	Precision	Recall	AUC
Random Forest	0.775 ± 0.161	0.798 ± 0.168	0.778 ± 0.207	0.767 ± 0.175	0.767 ± 0.175
Gradient Boosting	0.777 ± 0.161	0.802 ± 0.164	0.778 ± 0.204	0.764 ± 0.180	0.764 ± 0.180
AdaBoost	0.767 ± 0.161	0.783 ± 0.177	0.779 ± 0.209	0.757 ± 0.168	0.757 ± 0.168
SVM	0.709 ± 0.184	0.741 ± 0.200	0.714 ± 0.229	0.715 ± 0.165	0.715 ± 0.165
Decision Tree	0.671 ± 0.177	0.696 ± 0.193	0.720 ± 0.220	0.650 ± 0.203	0.650 ± 0.203
Multilayer Perceptron	0.649 ± 0.190	0.667 ± 0.227	0.721 ± 0.244	0.645 ± 0.191	0.645 ± 0.191
K-Nearest Neighbors	0.666 ± 0.164	0.722 ± 0.165	0.687 ± 0.213	0.642 ± 0.176	0.642 ± 0.176
Naive Bayes	0.600 ± 0.182	0.573 ± 0.236	0.722 ± 0.262	0.605 ± 0.175	0.605 ± 0.175
Majority	0.617 ± 0.232	0.734 ± 0.203	0.617 ± 0.231	0.500 ± 0.000	0.500 ± 0.000

Table 4. Average Performance Metrics Across Cross Validation Folds Using Deep ROC Analysis [43].

	Range	AUC	Precision	Recall	Specificity
Full Range	[0, 1]	0.816 ± 0.201	0.768 ± 0.178	0.816 ± 0.201	0.816 ± 0.201
High Risk (Low <i>FP</i>)	[0, 1/3]	0.796 ± 0.257	0.822 ± 0.267	0.651 ± 0.330	0.906 ± 0.221
Medium Risk (Medium <i>FP</i>)	[1/3, 2/3]	0.814 ± 0.217	0.688 ± 0.177	0.855 ± 0.214	0.245 ± 0.272
Low Risk (High <i>FP</i>)	[2/3, 1]	0.876 ± 0.249	0.620 ± 0.181	0.940 ± 0.145	0.016 ± 0.056

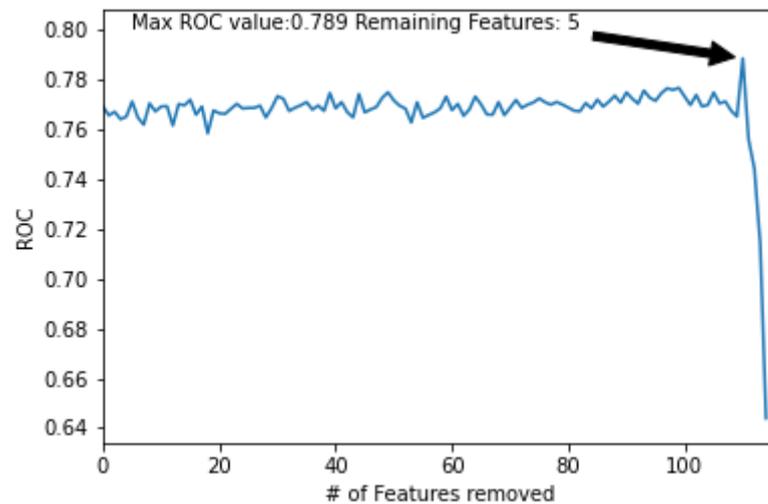
To improve the generalizability of the model and identify the most important features, we conducted feature selection while aiming to keep the AUC consistent. We used the feature importance function from the scikit-learn library; the output is a percentage representing the importance of each feature to a particular classifier. 116 features were reduced to 14 by the classifier, as shown in Table 5. We repeated the process of finding the AUC and then removing the least-important feature until only one feature remained. Figure 2 shows the different outcomes as more features were removed. We found no measurable dropoff in AUC until fewer than five features remained. The model with only these five features provided the highest AUC. Table 6 presents the ranked features from the classifier when only five remain. Those features were Gaze Points: Outside, Outside From Outside, Dwell Duration: Outside, Stimulus Duration, and Dwell Duration Average: Experience.

Table 5. Important Features for Random Forest.

Feature Name	Importance Percentage
Outside From Outside	8.258
Gaze Points: Outside	7.557
Number of Fixations: Outside	4.965
Dwell Duration Average: Outside	4.624
Dwell Duration: Outside	4.618
Stimulus Duration	2.290
Fractal Dimension Average	2.190
Number of Fixations: Experience	2.084
Dwell Rate: Education	2.068
Dwell Duration Average: Experience	1.914
Fractal Dimension	1.873
Dwell Rate: Experience	1.760
Dwell Duration: Experience	1.738
Dwell Rate: Outside	1.736

Table 6. Important Features for Random Forest with Optimal AUC.

Feature Name	Importance Percentage
Gaze Points: Outside	28.630
Outside From Outside	18.812
Dwell Duration: Outside	18.151
Stimulus Duration	18.110
Dwell Duration Average: Experience	16.297

**Figure 2.** The area under the ROC curve as features are removed.

5. Discussion

Though the goal was to discover what sections in a resume are the most important to recruiters, our results show that the most important features involve the Outside AOI. This outcome is peculiar because this area of the resumes is void of any actual content. We theorize that these features are capturing behavioral signals that indicate that the recruiter is thinking. When the recruiter looks away from the resume, they may be considering whether they have a good fit for the applicant. This hypothesis is bolstered by the fact that the highest-ranking feature unrelated to the Outside AOI is “Stimulus Duration”. The longer a recruiter observed a resume, the more likely they were to move that resume to the next level. This extra screening time would provide additional opportunities for recruiters to deliberate before reaching a decision on a resume. While this outcome does not give insight into how to design a resume, it does give applicants context for how to interpret a recruiter’s behavior. A recruiter looking away from the resume can be a good sign that they are engaging with the resume, not a signal that they dislike the resume.

Three of the five features selected in the model with optimal AUC were related to the Outside AOI. The Gaze Points feature captured how many samples were taken in the Outside AOI, without any context of where the gaze transitioned from or how long the gaze stayed there. The Outside to Outside Transitions feature captured how many times that fixations in the Outside AOI were followed by another fixation in the Outside AOI. The Dwell Duration feature captured how long the gaze stayed in the Outside AOI. Naturally, these features would be correlated, but each focuses on a different type of attention. Gaze Points can be high while the other two features are low if the recruiter frequently looks Outside for an instant before moving back to other AOIs. This behavior could occur when skimming the resume, where the recruiter’s gaze is jumping all over instead of fixating on a specific area to think about the resume. To that point, a general theme of the top five features is that the longer the recruiter spends reviewing a resume, the more likely they are to move it on to the next phase.

It is important to note that the recruiter should be looking at specific parts of the resume during the review process to improve the chances of the resume moving on. The highest performing features are connected to the recruiter fixating and dwelling on the Outside AOI, Experience AOI, and the Education AOI to a lesser extent. In other words, the high Stimulus Duration occurs because the recruiter was reading or pausing and thinking, not simply skimming or continuously jumping from section to section. We hypothesize that the latter type of behavior indicates that the reviewer is searching for information to make a decision about the applicant, information they perhaps did not find initially. With that in mind, we recommend that sections such as Experience and Education have clear and concise descriptions of the applicant's background to make it easier for the recruiter to understand the applicant's skillset.

6. Limitations and Future Work

The results presented are impacted by two main limitations: recruitment scope and lack of actionable takeaways for resume formatting. This study focused on recruiters evaluating computer science resumes from one location. As such, our observations may not be representative of recruiter practices in other areas. To address the latter point, this study should be repeated in other geographic locations across the country.

The majority of the most important features for determining whether a resume moved on to the next level were not connected to specific AOIs. In other words, no insights are gained with respect to the AOI sections for skills, projects, introduction, and address. This does not necessarily indicate that these sections are not important to recruiters, only that the eye-tracking data from these sections were not substantially different between resumes moved to the next level and those that failed to do so. Additionally, this study did not track the sequence order used by recruiters as they screened resumes, so this is a behavioral data source that is currently unexplored in the analysis. Future work has the potential to build upon this study and uncover more about the resume screening process.

Based on the relationships between the most important features, we discuss hypotheses for what recruiter behaviors these features detect. In general, a resume moving to the next phase is correlated with the amount of time the recruiter spends reviewing the resume. Because time spent dwelling in the Outside AOI and Experience AOI is important, we believe that this outcome shows that how the recruiter is spending their time matters. We encourage future work to investigate how the recruiter's behavior is tied to the outcomes for the resume.

7. Conclusions

Resumes are an integral part of the hiring process for recruiters; however, to applicants, it is not always clear what section of the resume is most important to recruiters and why they were ultimately rejected from a position. In this work, through a combination of eye-tracking and machine learning, we aimed to understand where employers are looking to gain insight into what they are searching for on resumes. Specifically, we developed a machine learning pipeline that utilized features extracted from recruiter eye-tracking data and a Random Forest classifier to recognize when a recruiter would move a resume onto the next stage of recruiting or not, with an AUC of 0.767. When investigating the most important features, we found that features in which the recruiter looked outside of resume AOIs were the most informative towards understanding whether or not a resume would move on, followed by total time spent reviewing the resume. Our takeaway from this observation is that these features indicate that the recruiter is contemplating the resume and is thinking about whether the applicant is a good fit. Additionally, features extracted from both the Experience and Education AOIs also appeared among the most important features, with Experience, in particular, appearing four times. More specifically, longer view times in the Experience AOI were correlated with resumes moving to the next level. Based on this finding, job applicants should focus on these particular sections, providing clear and sufficient descriptions of past experiences.

Author Contributions: A.P. took the lead and did the bulk of the initial work on the project. C.P. helped with writing and additional analyses. J.C. improved the writing and came up with additional ideas for analyses. J.N.L. was PI on the NSF grant and came up with the idea for the paper with T.H. She and her Co-PIs, G.A. and T.H., also met regularly with the rest of the team to discuss the paper, make improvements, and comment on different iterations of the paper. The work came out of T.H.'s lab. All authors have read and agreed to the published version of the manuscript.

Funding: This material is based upon work supported by the National Science Foundation under Grant Numbers 1658758 and 1658760. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Data Availability Statement: The data presented in this study on FigShare at doi 10.6084/m9.figshare.23582856

Conflicts of Interest: There is no conflict of interest with this work.

References

- Hornsby, J.S.; Smith, B.N. Resume content: What should be included and excluded. *SAM Adv. Manag. J.* **1995**, *60*, 4.
- Thoms, P.; McMasters, R.; Roberts, M.R.; Dombkowski, D.A. Resume Characteristics as Predictors of an Invitation to Interview. *J. Bus. Psychol.* **1999**, *13*, 339–356. doi:10.2297/4232557 [CrossRef]
- Kulkarni, S.B.; Che, X. Intelligent software tools for recruiting. *J. Int. Technol. Inf. Manag.* **2019**, *28*, 2–16. [CrossRef]
- Shellenbarger, S. Life & Arts: Make Your Job Application Robot-Proof—It Takes Planning to Make Sure AI Gatekeepers Don't Bounce Your Resume before a Human Can Make a Call. Available online: <https://www.wsj.com/articles/make-your-job-application-robot-proof-11576492201> (accessed on 10 November 2022)
- Schramm, R.M.; Dortch, N.R. An analysis of effective resume content, format, and appearance based on college recruiter perceptions. *Bull. Assoc. Bus. Commun.* **1991**, *54*, 18–23. [CrossRef]
- Noonan, R. *STEM Jobs: 2017 Update*; ESA Issue Brief# 02-17; US Department of Commerce: Washington, DC, USA, 2017. Available online: <https://eric.ed.gov/?id=ED594354> (accessed on 10 November 2022).
- Jobvite. Recruiting Benchmark Report. Available online: <https://www.jobvite.com/wp-content/uploads/2019/03/2019-Recruiting-Benchmark-Report.pdf> (accessed on 10 November 2022).
- Kabicher, S.; Motschnig-Pitrik, R.; Figl, K. 2009. What competences do employers, staff and students expect from a Computer Science graduate? In Proceedings of the 39th IEEE Frontiers in Education Conference, San Antonio, TX, USA, 18–21 October 2009; pp. 1–6. [CrossRef]
- Risavy, S. The Resume Research Literature: Where Have We Been and Where Should We Go Next? *J. Educ. Dev. Psychol.* **2017**, *7*, 169. [CrossRef]
- Guo, S.; Hammond, T. *ResuMatcher: A Personalized Resume-Job Matching System*; Texas A & M University: College Station, TX, USA, 2015.
- Stephen, B.K. Impressions of the Resume: The Effects of Applicant Education, Experience, and Impression Management. *J. Bus. Psychol.* **1994**, *9*, 33.
- Brown, B.K.; Champion, M.A. Biodata phenomenology: Recruiters' perceptions and use of biographical information in resume screening. *J. Appl. Psychol.* **1994**, *79*, 897–908. [CrossRef]
- Bretz, J.R.D.; Rynes, S.L.; Gerhart, B. Recruiter Perceptions of Applicant Fit: Implications for Individual Career Preparation and Job Search Behavior. *J. Vocat. Behav.* **1993**, *43*, 310–327. [CrossRef]
- Cable, D.M.; Judge, T.A. Interviewers' Perceptions of Person-Organization Fit and Organizational Selection Decisions. *J. Appl. Psychol.* **1997**, *82*, 546. [CrossRef]
- Roth, P.L.; Bobko, P. College grade point average as a personnel selection device: Ethnic group differences and potential adverse impact. *J. Appl. Psychol.* **2000**, *85*, 399–406. [CrossRef]
- Barr, T.F.; McNeilly, K.M. The value of students' classroom experiences from the eyes of the recruiter: Information, implications, and recommendations for marketing educators. *J. Mark. Educ.* **2002**, *24*, 168–173. [CrossRef]
- Adkins, C.L.; Russell, C.J.; Werbel, J.D. Judgments of fit in the selection process: The role of work value congruence. *Pers. Psychol.* **1994**, *47*, 605–623. [CrossRef]
- Clark, J.G.; Walz, D.B.; Wynekoop, J.L. Identifying exceptional application software developers: A comparison of students and professionals. *Commun. Assoc. Inf. Syst.* **2003**, *11*, 8. [CrossRef]
- Roth, P.L.; BeVier, C.A.; Switzer, F.S., III; Schippmann, J.S. Meta-analyzing the relationship between grades and job performance. *J. Appl. Psychol.* **1996**, *81*, 548. [CrossRef]
- Hutchinson, K.L.; Brefka, D.S. Personnel Administrators' Preferences for Résumé Content: Ten Years After. *Bus. Commun. Q.* **1997**, *60*, 67–75. [CrossRef]
- Quiñones, M.A.; Ford, J.K.; Teachout, M.S. The relationship between work experience and job performance: A conceptual and meta-analytic review. *Pers. Psychol.* **1995**, *48*, 887–910. [CrossRef]

22. Singer, M.S.; Bruhns, C. Relative effect of applicant work experience and academic qualification on selection interview decisions: A study of between-sample generalizability. *J. Appl. Psychol.* **1991**, *76*, 550. [[CrossRef](#)]
23. Hutchinson, K.L. Personnel Administrators' Preferences for Résumé Content: A Survey and Review of Empirically Based Conclusions. *J. Bus. Commun.* (1973) **1984**, *21*, 5–14. [[CrossRef](#)]
24. Cole, M.S.; Rubin, R.S.; Feild, H.S.; Giles, W.F. Recruiters' Perceptions and Use of Applicant Résumé Information: Screening the Recent Graduate. *Appl. Psychol.* **2007**, *56*, 319–343. [[CrossRef](#)]
25. Sulastri, A.; Handoko, M.; Janssens, J.M.A.M. Grade point average and biographical data in personal resumes: predictors of finding employment. *Int. J. Adolesc. Youth* **2015**, *20*, 306–316. [[CrossRef](#)]
26. Baert, S.; Neyt, B.; Siedler, T.; Tobback, I.; Verhaest, D. Student internships and employment opportunities after graduation: A field experiment. *Econ. Educ. Rev.* **2019**, *83*, 1–11 [[CrossRef](#)]
27. Ferguson, E. Changing qualifications for entry-level application developers. *J. Comput. Sci. Coll.* **2005**, *20*, 106–111.
28. Aasheim, C.; Shropshire, J.; Li, L.; Kadlec, C. Knowledge and skill requirements for entry-level IT workers: A longitudinal study. *J. Inf. Syst. Educ.* **2019**, *23*, 8.
29. Howard, A. College experiences and managerial performance. *J. Appl. Psychol.* **1986**, *71*, 530–552. [[CrossRef](#)]
30. Nemanick, J.R.C.; Clark, E.M. The Differential Effects of Extracurricular Activities on Attributions in Résumé Evaluation. *Int. J. Sel. Assess.* **2002**, *10*, 206–217. [[CrossRef](#)]
31. Rubin, R.S.; Bommer, W.H.; Baldwin, T.T. Using extracurricular activity as an indicator of interpersonal skill: Prudent evaluation or recruiting malpractice? *Hum. Resour. Manag.* **2002**, *41*, 441–454. [[CrossRef](#)]
32. Poole, A.; Ball, L.J. Eye tracking in HCI and usability research. In *Encyclopedia of Human Computer Interaction*; IGI Global: Hershey, PA, USA 2006; pp. 211–219. [[CrossRef](#)]
33. Mele, M.L.; Federici, S. Gaze and eye-tracking solutions for psychological research. *Cogn. Process.* **2012**, *13*, 261–265. [[CrossRef](#)]
34. El Haj, M.; Lenoble, Q. Eying the future: Eye movement in past and future thinking. *Cortex* **2018**, *105*, 97–103. [[CrossRef](#)]
35. Diaz, C.S. Updating Best Practices: Applying On-Screen Reading Strategies to Résumé Writing. *Bus. Commun. Q.* **2013**, *76*, 427–445. [[CrossRef](#)]
36. Lahey, J.N.; Oxley, D.R. *Discrimination at the Intersection of Age, Race, and Gender: Evidence from a Lab-in-the-Field Experiment*; National Bureau of Economic Research Working Paper Series, No. 25357; National Bureau of Economic Research: Cambridge, MA, USA, 2018. [[CrossRef](#)]
37. Shalev-Shwartz, S.; Ben-David, S. *Understanding Machine Learning: From Theory to Algorithms*; Cambridge University Press: New York, NY, USA 2014.
38. Yu, K.; Guan, G.; Zhou, M. Resume information extraction with cascaded hybrid model. In Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL'05), Ann Arbor, MI, USA, 25–30 June 2005; pp. 499–506.
39. Guo, S.; Alamudun, F.; Hammond, T. Résumatcher: A personalized résumé-job matching system. *Expert Syst. Appl.* **2016**, *60*, 169–182. [[CrossRef](#)]
40. Roy, P.K.; Chowdhary, S.S.; Bhatia, R. A Machine Learning approach for automation of Resume Recommendation system. *Procedia Comput. Sci.* **2020**, *167*, 2318–2327. [[CrossRef](#)]
41. Alamudun, F.; Yoon, H.J.; Hudson, K.B.; Morin-Ducote, G.; Hammond, T.; Tourassi, G.D. Fractal analysis of visual search activity for mass detection during mammographic screening. *Med. Phys.* **2017**, *44*, 832–846. [[CrossRef](#)] [[PubMed](#)]
42. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
43. Carrington, A.; Manuel, D.; Fieguth, P.; Ramsay, T.; Osmani, V.; Wernly, B.; Bennett, C.; Hawken, S.; Magwood, O.; Sheikh, Y.; et al. Deep ROC analysis and AUC as balanced average accuracy, for improved classifier selection, audit and explanation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, *60*, 329–341. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.