Advance Sustainable Science, Engineering and Technology (ASSET)
Vol. 8, No.2, April 2026, pp. 026020230-01 ~ 02602030-017
ISSN: 2715-4211 DOI: https://doi.org/10.26877/asset.v8i2.3123

A Decision Support System Based on Transformer-Driven Sentiment Analysis of Social Media Data

Arie Christian Wibisono*, Elfindah Princes

[1]Master of Information Systems Management, Information System Managemnt Department, Bina Nusantara University, Jakarta, Indonesia.

*ariew383@gmail.com

Abstract. The growing availability of social media data offers new opportunities for decision support systems (DSS) in large-scale human resource screening. This study proposes a technology-driven DSS architecture integrating transformer-based sentiment analysis to support early-stage candidate profiling. Its novelty lies in combining IndoBERT-based sentence embeddings with a structured DSS layer that aggregates tweet-level sentiment into risk-aware recommendations, rather than treating sentiment classification as a standalone output. Using a quantitative experimental design, 5,000 public posts from 100 users were processed through an NLP pipeline incorporating mean-pooled embeddings, feature engineering, principal component analysis, and Support Vector Machine classification. The model achieved 69.1% accuracy, with weighted precision, recall, and F1-score of 0.694, 0.691, and 0.691, outperforming baseline models by 6.5–15.0 percentage points. Sentiment outputs are treated as probabilistic behavioral signals within an advisory DSS framework, not direct indicators of candidate suitability. Preliminary validation on 50 cases showed moderate correlations (ρ = 0.52–0.61) with conventional assessments. The system remains non-automated, incorporating confidence thresholds, uncertainty handling, and mandatory human oversight. Limitations include moderate accuracy, reliance on text-only data, and linguistic ambiguity.

Keywords: Decision support systems, transformer-based embeddings, social media sentiment modeling, NLP pipeline architecture, human-in-the-loop AI

(Received 2025-12-08, Revised 2026-04-03, Accepted 2026-04-27, Available Online by 2026-04-30)

1. Introduction

Over the past two decades, advances in information and communication technologies have enabled large-scale collection of unstructured digital data, particularly from social media platforms [1]. From an engineering perspective, this development raises a fundamental challenge: how unstructured, noisy, and context-dependent textual data can be transformed into reliable, structured indicators that support decision-making processes. Decision Support Systems (DSS) provide a formal framework for integrating data acquisition, analytical modeling, and human judgment, especially in domains requiring scalability, transparency, and accountability.

In the context of information systems, the utilization of social media data for recruitment purposes demonstrates a convergence of big data technology, predictive analytics, and advanced systems such as Decision Support Systems (DSS) [2]. For instance, 94% of recruiters currently use or plan to use social media for recruitment [3]. Moreover, the im- plementation of big data analytics in recruitment practices has shown a significant impact on transforming traditional selection processes. This confirms that implementing social media strategies in police recruitment has the potential to enhance transparency, accelerate information dissemination, and attract the interest of the digital generation to join the institution. An example highlighted by [3] is the success of the New Zealand Police, which experienced an 800% increase in recruitment website traffic after utilizing a social media campaign.

The importance of social media analysis for recruitment assessment is supported by previous research validating the effectiveness of this method. A Research [4] affirms that assessing candidates' psychological characteristics has been a cornerstone of recruitment for decades, and machine learning technology now enables accurate personality prediction based on digital content. This is further reinforced by another study [5], which explains that traditional personality assessments have numerous weaknesses compared to social media analysis. This approach offers a new paradigm by evaluating a user's social media account content to obtain more relevant results. Thus, social media analysis for recruitment assessment is not merely a technological trend but an evolution in methodology supported by strong empirical evidence from various inter- disciplinary studies. Despite the global trends and proven po- tential, the recruitment process within the Indonesian National Police (Polri) remains largely conventional [6]. An analysis of the 2021 Polri Civil Service Candidate Recruitment data reveals several significant problems:

1) Inefficient Process Duration: The process is exception- ally lengthy, taking 5-6 months from registration to the announcement of results to select only 100 positions [7].
2) Limitations of Conventional Profiling: The current profiling system relies solely on document verification, academic tests, and interviews without leveraging avail- able digital data [8].
3) Declining Applicant Interest: A concerning trend shows a 28.4% decrease in interest over three years, with the number of applicants dropping from 13,125 in 2022 to 9,385 in 2025 [9].
4) Administrative Bottlenecks: The current administrative verification process requires participants to be physically present at local police headquarters (Panda/Polda) with over 15 different documents, leading to inefficiencies and susceptibility to human error [10].

Despite extensive research on sentiment analysis and personality prediction from social media, most existing studies focus on improving classification accuracy as an isolated objective. Limited attention has been given to how sentiment outputs can be systematically integrated into a decision support architecture that translates low-level predictions into interpretable, risk-aware indicators for human decision-makers. In particular, there is a lack of engineering-oriented frameworks that address aggregation across multiple texts, uncertainty handling, confidence calibration, and the explicit separation between predictive signals and final decisions. This gap limits the practical applicability of sentiment analysis in high-stakes decision contexts [11-14]. To address these challenges, this research aims to develop a Decision Support System (DSS) that integrates social media analysis with human resource information systems. Social media presents a valuable data source, as a survey conducted for this study (see Fig. 1) indicates that over 80% of newly recruited police officers are active on Twitter/X [15-18].

However, utilizing social media data presents its own set of complex challenges. Social media

generates unstructured big data in diverse formats, requiring an information systems approach capable of processing, analyzing, and interpreting it. This involves tackling challenges in data preprocessing, feature extraction, pattern recognition, and machine learning. Further- more, critical issues of information security and data privacy, particularly compliance with regulations like Indonesia's Personal Data Protection Law, must be addressed. Additionally, the inherent complexity and ambiguity of sentiment analysis on informal Indonesian social media text, characterized by code-mixing, slang, sarcasm, and contextual nuances, impose fundamental limitations on classification accuracy [19].

This study proposes a Decision Support System (DSS) that transforms unstructured social media text into structured decision indicators using transformer-based semantic modeling, feature aggregation, and uncertainty-aware risk scoring. The system generates probabilistic sentiment profiles to support, not replace, human judgment and is positioned as a supplementary screening tool that assists assessors in identifying potential indicators during early evaluation. Human oversight remains mandatory to ensure responsible use, while the underlying framework is designed to be transferable to other assessment contexts requiring hybrid human–AI decision support [20].

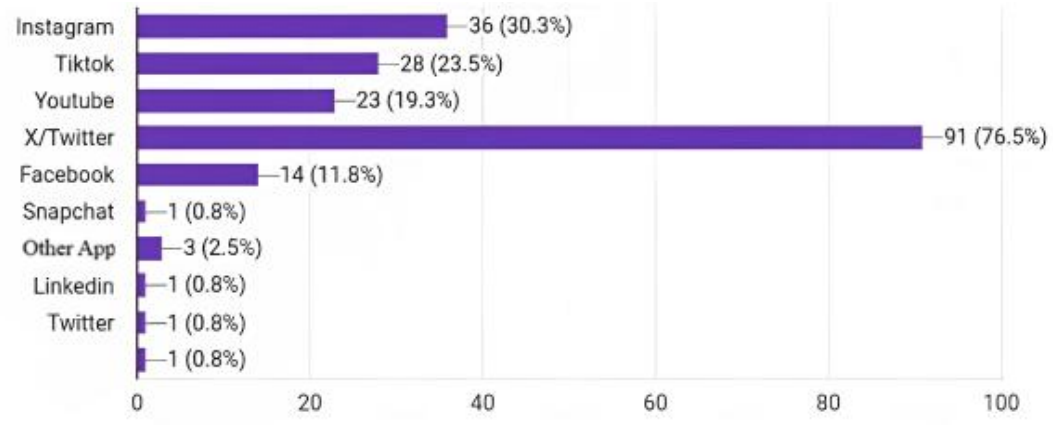

Figure. 1. Results of a Survey on Social Media Use Among New Police Officers 2024-2025.

The contributions of this study are threefold. First, it proposes an end-to-end engineering architecture for integrating transformer-based sentiment analysis into a decision support system, explicitly addressing aggregation, uncertainty, and human oversight. Second, it demonstrates how unstructured social media text can be converted into structured, interpretable risk indicators suitable for high-stakes screening contexts, supported by quantitative evaluation and preliminary validation. Third, it introduces a governance-aware DSS design that enforces advisory-only usage through confidence thresholds and human-in-the-loop mechanisms, contributing to responsible AI deployment within decision support research.

2. Method

2.1. Research Paradigm and Design

This study employs a positivist paradigm [21]with a quan- titative approach, in which social reality can be objectively measured through empirical and statistical methods. The theo- retical framework adopts digital behavior analysis and compu- tational personality psychology [22] as the foundation for understanding personality manifestations in a digital context. The research design adopts a quantitative experimental approach with a digital case study inspired by [23], namely modeling to simulate the sentiment analysis process based on the social media data of prospective Indonesian National Police (Polri) members. A computational mixed-method approach combines Natural Language Processing (NLP) techniques and behavior analysis to analyze users' social media digital footprints [24].

Methodologically, the research implements a supervised machine learning approach for sentiment classification using Support Vector Machine (SVM) with an RBF kernel and Random Forest as proposed methods, compared against base- line classifiers including Naive Bayes, Logistic Regression, and Decision Tree. The CRISP-DM framework is adopted as the methodological foundation to ensure a systematic process from data understanding to the presentation of findings. An
illustration of the research design is shown in Fig. 2.

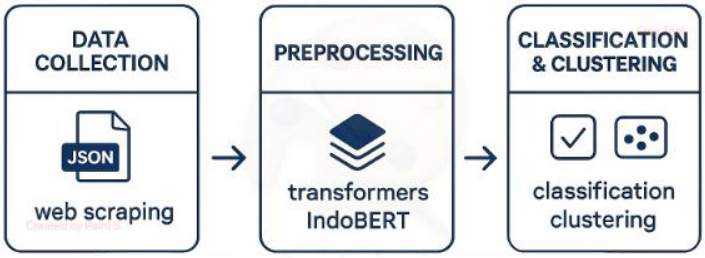

Figure 2. Research Design Illustration.

The analysis pipeline was built in Python using the trans- formers, scikit-learn, and TensorFlow libraries. The pipeline consists of three main modules:

1) Text Processing Pipeline: Text preprocessing includes cleaning non-alphabetic elements, case normalization, tokenization, and handling local language and specific Polri terminology.
2) Feature Extraction Module: Feature extraction uses a hybrid approach combining semantic representation with statistical features. Semantic features are extracted using the pre-trained IndoBERT-base-p1 model [25]. Specif- ically, tokenized input sequences are passed through the 12-layer transformer to obtain contextualized token embeddings from the final hidden layer. Rather than using only the [CLS] token representation, we apply mean pooling across all token embeddings (excluding padding tokens) to produce a single 768-dimensional sentence-level embedding vector. This mean pooling approach captures information from the entire input sequence, providing a more comprehensive semantic representation than relying solely on the [CLS] token. The mathematical formulation is:

$$hsentence = \frac{1}{n}\sum \ {}^{n}_{i=1}hi \qquad (1)$$

where hi represents the hidden state of the i-th token from IndoBERT's final layer, n is the number of non- padding tokens, and hsentence ∈ R768 is the resulting sentence embedding.
3) Machine Learning Module: The extracted feature vec- tors (58 dimensions: 50 from PCA-reduced IndoBERT embeddings + 8 from lexical/engagement features) are fed into classification models. The primary model is a Support Vector Machine (SVM) with Radial Basis Function (RBF) kernel, chosen for its effectiveness in handling high-dimensional, non-linear feature spaces. Hyperparameter optimization was conducted using grid search with 5-fold stratified cross-validation on the training set. For the SVM with RBF kernel, the regularization parameter C ∈ {0.1, 1, 10, 100} and kernel width γ ∈ {0.001, 0.01, 0.1, 1} were evaluated. Model selection was based on mean cross-validation F1-score to balance precision and recall across sentiment classes. The final configuration (C = 10, γ = 0.01) demonstrated the most stable performance with minimal variance across folds, reducing overfitting risk.

The combination of transformer-based semantic embeddings and handcrafted lexical and engagement features is motivated by complementary representational strengths. While transformer embeddings capture deep contextual semantics, they do not explicitly encode structural properties such as message length, punctuation density, or social engagement signals. Handcrafted features provide interpretable indicators of expressive intensity and interaction behavior, which are relevant for sentiment manifestation in social media. Prior studies have shown that hybrid feature representations often outperform purely neural embeddings in noisy, informal text environments. This design choice aims to

enhance robustness rather than maximize raw classification accuracy alone.

2.2. Population, Sample, and Data Collection

The research population consists of individuals aged 18-35 who are active on Twitter with public profiles, representing the demographic characteristics of prospective Polri members in the West Java region. The sample selection uses a purposive sampling technique with the following criteria: (1) public accounts without privacy restrictions, (2) a minimum of 50 posts in the last 6 months, (3) diversity of content (opinions, social interactions, personal narratives), (4) consistent activity (posting at least 2 times/week), and (5) Indonesian as the primary language of communication. Sampling stratification considers behavioral diversity (activity level, interaction pat- terns) and content type (personal, opinion, social content) to ensure representative coverage of different social media usage patterns.

Data was collected through web scraping using Python scripts with the snscrape library for Twitter/X and Sele- nium for platforms with dynamic content. The selection of Twitter/X as the primary platform is based on its adoption rate among the target demographic (>80% active), its text- centric characteristics, and its high level of expressiveness for personal opinions. The collected data includes post text, profile descriptions, comments, metadata, and social engagement pat- terns. To ensure compliance with privacy regulations (GDPR, Indonesian Personal Data Protection Law), only public data was collected and all data was anonymized without including explicit personal identifiers. Data is stored in JSON/CSV format using SQLite or MongoDB.

Sentiment labels were assigned at the tweet level using a semi-automatic approach, with initial polarity generated by a pre-trained Indonesian lexicon-based classifier and refined by two independent annotators following predefined guidelines. Disagreements were resolved through consensus, yielding a Cohen's Kappa of 0.78. To prevent temporal data leakage, tweets were aggregated at the user level prior to train–test splitting, ensuring all posts from the same user appeared in only one split. While the dataset spans 2022–2025, temporal effects were not explicitly modeled and are acknowledged as a limitation.

2.3. Research Variables and Instruments

The study examines the relationship between independent variables (digital features) extracted from two categories: (1) Linguistic Features include semantic representation using In- doBERT to understand hidden meanings, sentiment analysis to measure emotional polarity, and language complexity metrics (readability, vocabulary variation) that reflect cognitive abil- ity; (2) Digital Behavioral Features include activity patterns (posting frequency, active time distribution), social networking metrics (centrality, reciprocity), and engagement level as a proxy for social involvement. Validation is performed by triangulating with conventional psychometric test results to ensure the validity of the findings.

2.4. Data Analysis Techniques

Data analysis is conducted in stages integrating exploration, transformation, and modeling techniques:

- Exploratory Data Analysis (EDA Descriptive statistics are used to examine text length, word count, engagement metrics (likes, retweets, replies), and sentiment class distribution to assess data characteristics and class balance
- Feature Engineering and Dimensionality Reduction: Semantic features are extracted using IndoBERT-base-p1, producing 768-dimensional embeddings via mean pooling. PCA is applied to retain 95% of variance, reducing dimensionality to approximately 50–100 components. These are combined with lexical features (text length, word count, sentiment scores) and engagement metrics, then normalized using StandardScaler.
- Modeling and Validation: Model performance is evaluated using accuracy, precision, recall, F1-score, Matthews Correlation Coefficient (MCC), and Cohen's Kappa, supported by AUC-ROC curves and confusion matrices. Five-fold stratified cross-validation is employed, with baseline models compared against SVM (RBF kernel) and Random Forest, and results reviewed with domain experts [26].

An ablation study was conducted to justify architectural and feature choices by evaluating four configurations: transformer embeddings only, handcrafted features only, transformer embeddings without PCA, and the full hybrid model, using accuracy and weighted F1-score under identical cross-validation. Results show that the hybrid model outperforms single-component configurations, with handcrafted features enhancing semantic embeddings and PCA improving stability. Error analysis reveals that misclassifications mainly occur between Neutral and weakly Positive or Negative tweets, particularly in short, sarcastic, or code-mixed texts, indicating intrinsic linguistic ambiguity rather than model instability.

2.5 Ethical Governance and Data Management

This study exclusively utilized publicly accessible social media data. Data collection was conducted only after obtaining informed consent from participants whose accounts were analyzed. All personal identifiers were removed during preprocessing, and user IDs were anonymized using irreversible hashing. Data storage followed a principle of minimal retention, with raw data access restricted to authorized researchers. The system complies with applicable data protection regulations, including the Indonesian Personal Data Protection Law and GDPR principles. Data were encrypted at rest and deleted after analysis completion.

3. **Result and Discussion**

3.1. Feature Extraction and Preprocessing Pipeline

Raw tweet text was tokenized using the IndoBERT tokenizer with a maximum sequence length of 128 and processed by the pre-trained IndoBERT-base-p1 model (12 layers, 768 dimensions). Semantic representations were obtained using mean pooling over the final-layer hidden states of all non-padding tokens, rather than relying on the [CLS] token, to better capture distributed semantic information in informal text [27].

The resulting 768-dimensional semantic embeddings were then augmented with handcrafted features: (1) lexical features including text length (mean: 112.4 characters), word count (mean: 16.1 words), and punctuation density; (2) engagement metrics including like count (mean: 11.8), retweet count (mean: 3.2), and reply count (mean: 2.7); and (3) sentiment polarity scores from a rule-based Indonesian sentiment ana- lyzer. All features were standardized using StandardScaler. To manage dimensionality, PCA was applied to the IndoBERT embeddings, reducing them from 768 to 50 dimensions while retaining 95% of explained variance. The final feature matrix combining reduced semantic features (50-dim) with lexical and engagement features (8-dim) resulted in a 58-dimensional representation per tweet.

From a systems perspective, computational efficiency is critical for large-scale screening scenarios. Feature extraction using IndoBERT represents the primary computational cost; however, embedding generation is performed offline and cached, enabling scalable deployment. Once embeddings are generated, classification and risk aggregation operate with low latency, allowing near-real-time profile updates. The modular pipeline design supports horizontal scaling and batch processing, making the system suitable for high-volume applicant pools without compromising interpretability or governance constraints.

3.2. Dataset Characteristics

The final dataset comprised 5,000 Indonesian tweets col- lected from 100 unique users over the period 2022-2025. The data exhibited a relatively balanced distribution across sentiment classes: Neutral (38.3%, n=1,916), Positive (33.3%, n=1,665), and Negative (28.4%, n=1,419). This near-balanced distribution is favorable for classification tasks as it mitigates class imbalance issues that could bias model performance toward majority classes. The dataset was partitioned using an 80:20 train-test split (4,000 training samples, 1,000 test sam- ples) with stratification to preserve class proportions, ensuring representative sampling across all sentiment categories.

The balanced class distribution establishes important perfor- mance baselines for interpretation: random chance would yield.

Table 1. Model Performance Comparison

Model	Category	Test Accuracy	Cross-Validation Mean	Cross-Validation Std. Dev.
Support Vector Machine (SVM)	Proposed	0.691	0.674	0.008
Random Forest	Proposed	0.640	0.646	0.012
Logistic Regression	Baseline	0.626	0.637	0.011
Naive Bayes	Baseline	0.576	0.570	0.004
Decision Tree	Baseline	0.541	0.516	0.017

Reference Baselines: Random Guessing = 33.3%, Majority Class = 38.3%

33.3% accuracy (assuming uniform guessing), while a naive majority-class classifier would achieve 38.3% accuracy by always predicting Neutral sentiment. Any meaningful model must substantially exceed these baselines to demonstrate gen- uine learning capability.

3.3. Classification Performance and Model Comparison

Table 1 presents the comparative performance of all evalu- ated models on the test set, along with their cross-validation scores. The proposed methods (Support Vector Machine and Random Forest) demonstrated superior performance compared to all baseline approaches.

While the proposed SVM model achieved the highest accuracy among the evaluated approaches (69.1%), this performance should be interpreted cautiously in the context of high-stakes decision support. Although the accuracy substantially exceeds random guessing (33.3%) and the majority-class baseline (38.3%), it remains within a moderate range that does not justify autonomous or deployment-ready decision-making. Instead, the reported performance indicates that the model is capable of extracting meaningful sentiment signals from noisy social media text, but with a non-negligible level of uncertainty that must be explicitly managed through system design and human oversight.

Random Forest, the second proposed method, attained 64.0% test accuracy with a cross-validation mean of 64.6% (± 1.2%). Among the baseline methods, Logistic Regression performed best with 62.6% accuracy, followed by Naive Bayes (57.6%) and Decision Tree (54.1%).

The proposed SVM model outperformed the best baseline method (Logistic Regression) by 6.5 percentage points, repre- senting a relative improvement of 10.4%. This improvement is statistically meaningful given the relatively low standard deviations observed in cross-validation (ranging from 0.4% to 1.7%), indicating stable and reliable model performance. The cross-validation results demonstrate good generalization across all models, with the SVM exhibiting exceptional stability (CV std = 0.8%), suggesting robust performance across different training subsets. In contrast, the Decision Tree baseline shows the highest variance (1.7%), indicating sensitivity to training data composition and a tendency toward overfitting.

Table 2. Comprehensive Svm Performance Metrics

Metric	Test Set	Cross-Validation
Accuracy	0.691	0.674 ± 0.008
Precision (Weighted)	0.694	0.676 ± 0.007
Recall (Weighted)	0.691	0.674 ± 0.008
F1-Score (Weighted)	0.691	0.675 ± 0.007
Precision (Macro)	0.686	-

Recall (Macro)	0.688	-
F1-Score (Macro)	0.686	-
Matthews Correlation Coefficient (MCC)	0.536	-
Cohen's Kappa	0.535	-

From a reliability perspective, the alignment between weighted and macro-averaged metrics Moderate MCC (0.536) and Cohen's Kappa (0.535) indicate varying prediction confidence across samples, supporting the use of uncertainty-aware decision support rather than deterministic judgments. Error analysis shows that misclassifications mainly occur between Neutral and weakly Positive or Negative sentiments, particularly in short, sarcastic, or code-mixed texts, reflecting intrinsic linguistic ambiguity and justifying confidence-based flagging and human review.

3.4. Detailed SVM Performance Analysis

Table 2 presents comprehensive performance metrics for the best-performing SVM model, demonstrating balanced per- formance across multiple evaluation criteria. The SVM model demonstrates strong and consistent performance across mul- tiple metrics. The weighted precision (0.694), recall (0.691), and F1-score (0.691) are nearly identical, indicating balanced performance without bias toward any particular class. The close alignment between macro-averaged metrics (precision: 0.686, recall: 0.688, F1: 0.686) and weighted metrics suggests that the model performs consistently across all sentiment classes, despite minor class imbalances in the dataset.

The Matthews Correlation Coefficient (MCC) of 0.536 and Cohen's Kappa of 0.535 provide robust measures of classifi- cation quality that are less sensitive to class imbalance than accuracy alone. Both values fall within the "moderate agree- ment" range (0.4-0.6), confirming that the model's predictions are substantially better than chance while acknowledging room for improvement. The near-identical values of MCC and Kappa (differing by only 0.001) indicate consistent agreement between these complementary measures of classification per- formance.

The minimal standard deviations in cross-validation results (all $< 0.8\%$) demonstrate exceptional model stability. The close alignment between test set performance (0.691) and cross-validation mean (0.674) suggests good generalization with minimal overfitting. The small gap of 1.7 percentage points between these values is well within acceptable limits and indicates that the model will likely perform consistently on new, unseen Indonesian social media data.

The balanced precision-recall performance is particularly noteworthy in the context of sentiment analysis applications. High precision minimizes false positives, crucial when the sys- tem is used for decision-making in sensitive contexts such as personnel assessment. Simultaneously, the comparable recall ensures that the model does not miss significant sentiment indicators. This balance makes the model suitable for practical deployment where both types of errors carry consequences.

3.5. ROC Curve Analysis and Class-Specific Performance

Figure 3 presents the Receiver Operating Characteristic (ROC) curves for the SVM classifier across all three sentiment classes. The Area Under the Curve (AUC) values demonstrate strong discriminative ability for all classes. The ROC analysis reveals excellent classification performance across all senti- ment classes. The Neutral class achieved the highest AUC of 0.881, indicating superior ability to distinguish neutral tweets from other sentiment categories. This strong performance can be attributed to the distinctive linguistic characteristics of neutral content, which typically employs more factual language, lower emotional intensity, and fewer sentiment- laden expressions compared to explicitly positive or negative tweets. Neutral statements often follow information-sharing patterns with minimal evaluative language, providing clear discriminative features for the classifier.

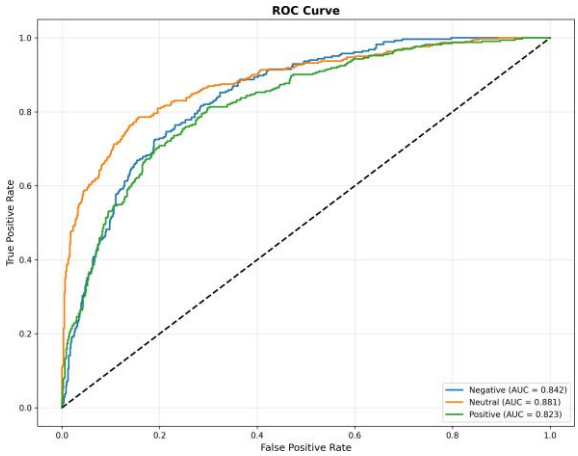

Figure 3. ROC curves for SVM classifier showing class-specific discrimination ability. All curves are substantially above the diagonal (random classifier), with AUC values ranging from 0.823 to 0.881.

The Negative class demonstrates strong discrimination (AUC = 0.842), driven by explicit negative markers such as *"kecewa,"* (disappointed) *"buruk,"* (bad), and "*parah*" (terrible)*,"* while the Positive class achieves slightly lower but robust performance (AUC = 0.823), reflecting the linguistic diversity of expressions like *"lumayan," "oke," "senang banget,"* and *"mantap."* Overall, balanced performance across sentiment classes (AUC range: 0.823–0.881; span = 0.058) indicates that the model does not disproportionately favor any single class and captures distinctive sentiment patterns effectively.

The evaluation results must be interpreted in conjunction with human–system interaction mechanisms. Rather than presenting sentiment scores as opaque outputs, the DSS exposes interpretable artifacts, including sentiment distributions, representative negative samples, and confidence levels. Recruiters interact with the system through a review dashboard that supports contextual interpretation and override capability. This interaction model reduces automation bias by ensuring that human judgment remains central, particularly in borderline or high-risk cases.

3.6. Decision Support System Framework: From Sentiment to Recommendation

The DSS framework transforms individual tweet-level sen- timent classifications into aggregated candidate profiles and actionable recommendations through a multi-stage process. This section details the technical implementation of converting sentiment analysis outputs into decision support metrics.

3.6.1. Sentiment Aggregation and Profile Generation

For each candidate with n tweets, the system computes an aggregate sentiment profile vector P = [ppos, pneu, pneg] representing the proportion of tweets in each sentiment category:

$$ps = \frac{1}{n}\sum \quad \frac{n}{i=1}1(yi = s) \qquad \text{for s} \in \{\text{pos, neu, neg}\} \quad (2)$$

where y*i* is the predicted sentiment label for tweet i, and 1(•) is the indicator function. Additionally, we compute sentiment consistency metrics to assess behavioral stability:

$$Entropy(P) = -\sum sps \qquad log2\,(ps) \quad (3)$$

Higher entropy (max = 1.585 bits for 3 classes) indicates inconsistent sentiment patterns, while lower entropy suggests more uniform behavior. We also track temporal sentiment trends by computing a 30-day rolling average of sentiment polarity to detect sudden behavioral shifts that may indicate concerning patterns.

3.6.2. Risk Scoring and Classification

The DSS assigns each candidate to one of four risk categories using a rule-based scoring system informed by the sentiment profile:

$$Risk\ Score = wneg \quad pneg \quad + went \quad Entropy(P) + weng \quad \hat{E}neg \qquad (4)$$

where the weights wneg = 0.5, went = 0.3, and weng = 0.2 were established through iterative refinement based on preliminary validation results and domain expert feedback. The weight hierarchy prioritizes direct negative sentiment presence (wneg) as the primary risk factor, informed by recruitment psychology literature indicating that expressed negativity correlates with occupational stress intolerance, interpersonal conflicts, and counterproductive work behaviors. Behavioral included to capture potential influence in spreading negativity. While these weights represent an informed heuristic rather than formally optimized parameters, the resulting risk scores demonstrated moderate correlation ($\rho = 0.61$, $p < 0.001$) with background check outcomes in preliminary validation (Table III). The integration of social media indicators in personnel assessment has gained empirical support [28], though requiring careful validation and ethical oversight. Future work should explore data-driven optimization of these weights as larger validation datasets become available.

Table 3. DSS Recommendation Logic

Risk Category	Sentiment Profile Condition	System Recommendation
Low Risk	p_pos > 0.4 and p_neg < 0.2	Proceed. Candidate shows positive digital behavior. Minimal additional screening required.
Moderate Risk	0.2 ≤ p_neg ≤ 0.4	Standard Review. Conduct routine background checks and psychological assessment.
High Risk	p_neg > 0.4 or Entropy > 1.3	Enhanced Review. Flag for in-depth interview and contextual review of negative posts.
Critical Risk	p_neg > 0.6 or violent/extremist indicators	Manual Investigation. Require full psychological evaluation and senior-level review.

E¯neg is the average Risk categories are defined as:

- Low Risk (Score < 0.25): pneg < 0.2 and ppos > 0.4: Predominantly positive profile
- Moderate Risk (0.25 ≤ Score < 0.5): Balanced senti- ment or moderate negative presence
- High Risk (0.5 ≤ Score < 0.7): pneg > 0.4 or high entropy with significant negative engagement
- Critical Risk (Score ≥ 0.7): pneg > 0.6 or extreme negativity patterns

The risk categorization scheme is based on exploratory analysis and treated as provisional heuristics, suitable for prioritization rather than automated decisions. Sensitivity analysis (±10% weight variation) shows stable Low and Critical categories, with Moderate and High risks exhibiting expected sensitivity. The DSS supports explainability through sentiment distributions, example outputs, and confidence scores, while governance rules mandate human review for all High and Critical risk cases.

3.6.3. Recommendation Generation
The system generates system structured recommendation for human reviewers using the following decisions logic.

3.6.4. Confidence Scoring and Uncertainty Handling
Given the model's 69.1% accuracy, the DSS incorporates uncertainty quantification to avoid overconfident recommendations. For each candidate, we compute a profile confidence score based on the SVM's decision function values:

$$Confidence = \frac{1}{n}\sum_{i=1}^{n} \frac{|f(xi)|}{max_j |f_j(xi)|} \quad (5)$$

where $f(x_i)$ is the SVM decision function value for tweet i, and the normalization ensures confidence $\in [0, 1]$. Candidates with low confidence scores (< 0.6) are automatically flagged for manual review regardless of risk category, acknowledging the model's limitations.

3.6.5. Integration with Existing Recruitment Workflow
Figure 4 illustrates how the DSS integrates with the conventional Polri recruitment process:

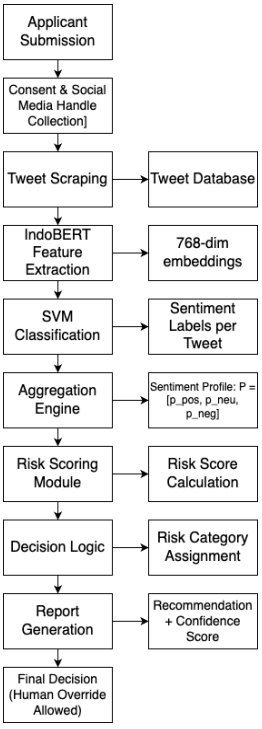

Figure. 4. DSS Integration Workflow. The system serves as an early screening layer, processing social media data in parallel with document verification to provide recruiters with candidate risk profiles before psychological testing.

The DSS operates as a parallel screening mechanism during the initial application phase:

1) Data Collection Phase: Upon receiving consent, ap- plicants' public social media handles are collected and validated.
2) Automated Profiling: The system scrapes recent tweets (past 6 months), applies the trained SVM classifier, and generates sentiment profiles.
3) Risk Assessment: Aggregated profiles are scored and categorized according to the risk matrix (Table III).

4) Recommendation Output: A structured report is gen- erated for each candidate, including:
   - Sentiment distribution visualization (pie chart)
   - Risk score and category
   - Confidence level and uncertainty flags
   - Top 5 most negative tweets (for High/Critical risk cases)
   - Recommended action (Proceed/Review/Investigate)
5) Human Review Integration: Recommendations are presented to recruitment officers via a dashboard in- terface. Officers can override system recommendations with documented justification. Final decisions remain with human assessors

Table 4. Correlation Between DSS Risk Scores and Recruitment Outcomes

Recruitment Outcome Metric	Spearman's ρ	p-value
Psychological Assessment Score	-0.58	< 0.001
Interview Panel Rating	-0.52	< 0.001
Background Check (Pass/Concern)	0.61	< 0.001

This scoring system enables systematic prioritization of re- viewer attention, focusing human expertise on high-risk cases while expediting low-risk candidates through the pipeline.

3.7. Validation of Social Media Sentiment as a Recruitment Indicator

The validation results presented in this section should be interpreted strictly as exploratory. The limited sample size (n = 50) and voluntary participation constrain statistical power and generalizability. Consequently, these findings do not establish predictive validity for operational deployment but rather provide preliminary evidence that motivates further large-scale and longitudinal evaluation. The pilot validation serves to assess directional consistency between DSS outputs and conventional assessments, not to confirm readiness for real-world adoption.

3.7.1. Validation Methodology

For these 50 candidates, we compared the DSS risk scores (generated from social media analysis) against three ground-truth metrics collected during conventional recruitment:

1) Psychological Assessment Scores: Standardized psy- chological tests administered by forensic psychologists, measuring traits such as emotional stability, stress toler- ance, and interpersonal skills (scale: 0-100).
2) Interview Panel Ratings: Averaged scores from three- member interview panels assessing candidate suitability, professionalism, and communication skills (scale: 1-5).
3) Background Check Findings: Binary classification (Pass/Concern) based on comprehensive background investigations including criminal records, employment history, and reference checks.

3.7.2. Correlation Analysis Results

Table IV presents Spear- man correlation coefficients between DSS risk scores and ground-truth metrics:

The negative correlations with psychological scores and interview ratings (higher DSS risk = lower scores) indicate that candidates flagged as high-risk by the system tend to receive lower evaluations in conventional assessments. The positive correlation with background check concerns (higher risk = more concerns) further supports predictive validity.

3.7.3. Contingency Table Analysis

Table V shows the agree- ment between DSS risk categories and final recruitment deci- sions:

The system correctly identified 12/22 (54.5%) of eventually rejected candidates as High/Critical risk, while 15/17 (88.2%) of Low-risk candidates were ultimately accepted. Cohen's Kappa = 0.47 indicates moderate agreement between DSS recommendations and final decisions.

Table 5. Dss Risk Category VS. Final Recruitment Decision (N=50)

DSS Risk	Accepted	Rejected	Total
Low Risk	15	2	17
Moderate Risk	10	8	18
High Risk	3	9	12
Critical Risk	0	3	3
Total	28	23	50

3.7.4. Limitations and Interpretations

These preliminary find- ings suggest moderate predictive validity, but several limita- tions warrant caution:

- Small sample size (n=50) limits generalizability
- Selection bias: Only candidates who consented to social media data collection were included
- Temporal confounds: Social media behavior may not reflect long-term personality traits
- Circularity risk: If recruiters were aware of DSS outputs (not the case here), this could artificially inflate correla- tions

Despite these limitations, the moderate-to-strong correla- tions (ρ = 0.52-0.61) provide initial empirical support for the hypothesis that social media sentiment patterns contain signal relevant to recruitment assessment. However, this validation is preliminary, and larger-scale longitudinal studies are essen- tial to establish robust predictive validity before operational deployment.

3.8. Performance Analysis and Challenges

While the SVM model achieved the highest accuracy (69.1%) among all evaluated approaches and substantially exceeded naive baselines (random: 33.3%, majority class: 38.3%), this performance level warrants careful interpretation in the context of high-stakes applications such as personnel recruitment. The balanced nature of the dataset (class distri- bution ranging from 28.4% to 38.3%) ensures that the reported accuracy reflects genuine classification capability rather than bias toward majority classes. However, the 69.1% accuracy, though representing a significant improvement over baseline models, reflects several fundamental challenges inherent to sentiment analysis on Indonesian social media text:

1) Linguistic Complexity: Indonesian social media ex- hibits extensive code-mixing (Indonesian-English), in- formal slang, regional dialects, and colloquial expres- sions. For example, tweets like "Finally paket dateng!" mix English with Indonesian, creating challenges for language models primarily trained on formal text. Jakarta-specific slang and regional variations further complicate accurate classification.
2) Contextual Ambiguity: Many tweets contain sarcasm, irony, or context-dependent sentiments that are difficult to disambiguate from text alone. Expressions like "Man- tap nih" can convey genuine satisfaction or sarcastic disappointment depending on context, requiring world knowledge beyond the available textual features.
3) Class Boundary Overlap: The distinction between Neutral and mildly Positive/Negative sentiments is in- herently fuzzy. Tweets with minimal sentiment markers create ambiguous cases where even human annotators might disagree, representing an intrinsic limitation of natural language rather than solely a model deficiency.
4) Multimodal Content Gap: Social media posts often rely on emojis, images, or external links to convey sentiment. Our text-only approach cannot capture these additional signals, potentially limiting accuracy for posts where non-textual elements carry critical sentiment in- formation.
5) Demographic Fairness Considerations: The current study does not include systematic bias analysis across demographic groups (gender, age, regional origin) due to privacy constraints and limited

demographic metadata in the dataset. This represents an important limitation, as algorithmic bias against certain demographics is a well- documented concern in ML-based recruitment systems [29], [30]. Potential bias sources include regional dialect variations, formality differences correlating with so- cioeconomic background, and inconsistent code-mixing treatment. Current safeguards include mandatory hu- man review, confidence-based flagging (threshold ¡ 0.6), explainability provisions showing flagged content for context, and strict advisory-only positioning. Future implementations must incorporate fairness auditing, in- cluding stratified performance analysis across protected attributes, adversarial debiasing techniques [31], and regular monitoring for disparate impact. The system should be deployed with explicit guidelines requiring human review to catch potential discriminatory patterns that automated systems might perpetuate.

Given these limitations, the proposed system should be positioned strictly as a decision support tool rather than an autonomous decision-making system. The 69.1% accuracy is sufficient for preliminary screening and flagging purposes, helping human reviewers prioritize cases and identify potential concerns, but insufficient for making final recruitment decisions without human oversight. The system's value lies in augmenting human judgment by processing large vol- umes of data efficiently, not in replacing expert assessment. Human reviewers must maintain final authority, using the system's outputs as one of multiple information sources in comprehensive candidate evaluation.

This positioning aligns with best practices in AI-assisted decision-making for high-stakes domains, where algorithmic systems serve to inform rather than determine outcomes. Regular auditing, validation against human assessments, and continuous monitoring for bias or drift are essential operational requirements for responsible deployment.

4. **Conclusion**

This study presents a technology-oriented Decision Support System (DSS) framework that integrates transformer-based sentiment analysis into an advisory screening architecture for large-scale candidate profiling. The primary technical contribution lies in the end-to-end system design that transforms unstructured social media text into structured, interpretable decision indicators through semantic embedding, feature aggregation, uncertainty-aware risk scoring, and human-in-the-loop governance. Rather than advancing sentiment classification accuracy in isolation, this work demonstrates how moderate-performing predictive models can be responsibly embedded within a DSS to support, but not replace, human judgment in high-stakes contexts.

Empirical evaluation shows that the proposed Support Vector Machine model achieves moderate accuracy (69.1%) and stable performance across multiple metrics, substantially exceeding naive baselines while remaining insufficient for autonomous decision-making. Accordingly, the system is explicitly positioned as an advisory tool that prioritizes transparency, confidence calibration, and mandatory human oversight. Although the experimental implementation is demonstrated within a specific application domain, the underlying architecture, NLP pipeline, and governance mechanisms are transferable to other decision support scenarios involving unstructured text, such as compliance monitoring, applicant pre-screening, or behavioral risk assessment.

Consistent with responsible AI principles, this work defines clear boundaries on system use: DSS outputs are probabilistic signals intended for prioritization and review allocation, not for exclusion or final decisions. Human reviewers retain full authority, supported by interpretable outputs and override mechanisms that mitigate automation bias.

This study is limited by moderate classification performance, dataset scope and potential sampling bias, an underpowered pilot validation (n = 50), and reliance on text-only analysis. Accordingly, the results demonstrate technical feasibility rather than deployment readiness. Future work should focus on domain-specific model fine-tuning, multimodal data integration, large-scale longitudinal validation, fairness auditing, and enhanced explainability to support responsible and accountable DSS deployment.

Declaration of AI and AI assisted technologies in the writing process
The authors declare that no artificial intelligence (AI) or AI-assisted technologies were used in the preparation, writing, or editing of this manuscript. All aspects of the work were conducted and written solely by the authors.

Declaration of Competing Interest
The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements
The authors would like to express sincere gratitude to all individuals and institutions who have supported the completion of this study, particularly the supervisors for their valuable guidance and feedback. Appreciation is also extended to colleagues, respondents, and the affiliated institution for their contributions and support. Finally, the authors would like to thank their family for their continuous encouragement and understanding throughout the research process.

References

[1] M. Syukri, "Situasi dan kondisi media digital di Indonesia: Meningkatkan keamanan dan ketahanan digital — The SMERU Research Institute," 2023. [Online]. Available: https://smeru.or.id/id/research-id/situasi-dan-kondisi-media-digital-di-indonesia-meningkatkan-keamanan-dan-ketahanan

[2] Z. Elmenzhi, S. Alji, A. Maghni, and M. Belamhitou, "Exploring the impact of big data analytics on recruitment practices," in *International Conference on Advanced Intelligent Systems for Sustainable Development*. Springer, 2024, pp. 882–898. https://doi.org/10.1007/978-3-031-91337-2_78

[3] S. Ziegler, "What every PD needs in a social media recruitment plan," Oct. 2020. [Online]. Available: https://www.police1.com/police-recruiting/articles/what-every-pd-needs-in-a-social-media-recruitment-plan-hSz8NU7u48RvtHYr/

[4] E. Grunenberg, H. Peters, M. J. Francis, M. D. Back, and S. C. Matz, "Machine learning in recruiting: Predicting personality from CVs and short text responses," *Frontiers in Social Psychology*, vol. 1, p. 1290295, 2024. https://doi.org/10.3389/frsps.2023.1290295

[5] R. Moraes, L. L. Pinto, M. Pilankar, and P. Rane, "Personality assessment using social media for hiring candidates," in *2020 3rd International Conference on Communication System, Computing and IT Applications (CSCITA)*. IEEE, 2020, pp. 192–197. https://doi.org/10.1109/CSCITA47329.2020.9137818.

[6] M. M. Tadesse, H. Lin, B. Xu, and L. Yang, "Personality predictions based on user behavior on the Facebook social media platform," *IEEE Access*, vol. 6, pp. 61 959–61 969, 2018. https://doi.org/10.1109/ACCESS.2018.2876502

[7] A. Souri, S. Hosseinpour, and A. M. Rahmani, "Personality classification based on profiles of social networks' users and the Five-Factor Model of personality," *Human-centric Computing and Information Sciences*, vol. 8, no. 1, p. 24, 2018. https://doi.org/10.1186/s13673-018-0147-4

[8] A. Naz, H. U. Khan, A. Bukhari, B. Alshemaimri, A. Daud, and M. Ramzan, "Machine and deep learning for personality traits detection: A comprehensive survey and open research challenges," *Artificial Intelligence Review*, vol. 58, no. 8, p. 239, 2025. https://doi.org/10.1007/s10462-025-11245-3

[9] H. Christian, D. Suhartono, A. Chowanda, and K. Z. Zamli, "Text-based personality prediction from multiple social media data sources using pre-trained language model and model averaging," *Journal of Big Data*, vol. 8, no. 1, p. 68, 2021. https://doi.org/10.1186/s40537-021-00459-1

[10] F. Habib, Z. Ali, A. Azam, K. Kamran, and F. M. Pasha, "Navigating pathways to automated personality prediction: A comparative study of small and medium language models," *Frontiers in Big Data*, vol. 7, p. 1387325, 2024. https://doi.org/10.3389/fdata.2024.1387325

[11] T. Joseph, "Natural language processing (NLP) for sentiment analysis in social media," *International Journal of Computing and Engineering*, vol. 6, no. 2, pp. 35–48, 2024.

[12] M. Ranjan, S. Tiwari, A. M. Sattar, and N. S. Tatkar, "A new approach for carrying out sentiment analysis of social media comments using natural language processing," *Engineering Proceedings*, vol. 59, no. 1, p. 181, 2024. **https://doi.org/10.3390/engproc2023059181**

[13] K. Chaudhary, M. Alam, M. S. Al-Rakhami, and A. Gumaei, "Machine learning-based mathematical modelling for prediction of social media consumer behavior using big data analytics," *Journal of Big Data*, vol. 8, no. 1, p. 73, 2021. https://doi.org/10.1186/s40537-021-00466-2

[14] S. B. Abkenar, M. H. Kashani, E. Mahdipour, and S. M. Jameii, "Big data analytics meets social media: A systematic review of techniques, open issues, and future directions," *Telematics and Informatics*, vol. 57, p. 101517, 2021. https://doi.org/10.1016/j.tele.2020.101517

[15] J. Gilbert, S. Hamid, I. A. T. Hashem, N. A. Ghani, and F. F. Boluwatife, "The rise of user profiling in social media: Review, challenges and future direction," *Social Network Analysis and Mining*, vol. 13, no. 1, p. 137, 2023. https://doi.org/10.1007/s13278-023-01146-0

[16] N. Han, S. Li, F. Huang, Y. Wen, Y. Su, L. Li, X. Liu, and T. Zhu, "How social media expression can reveal personality," *Frontiers in Psychiatry*, vol. 14, p. 1052844, 2023. https://doi.org/10.3389/fpsyt.2023.1052844

[17] M. D. Kamalesh and B. Bharathi, "Personality prediction model for social media using machine learning technique," *Computers and Electrical Engineering*, vol. 100, p. 107852, 2022. https://doi.org/10.1016/j.compeleceng.2022.107852

[18] Q. Tang, W. Jiang, Y. Du, and L. Lin, "An attention-based denoising framework for personality detection in social media texts," *arXiv preprint* arXiv:2311.09945, 2023. https://doi.org/10.1109/ISPA67752.2025.00097

[19] G. Alwafi and B. Fakieh, "A machine learning model to predict privacy-fatigued users from social media personalized advertisements," *Scientific Reports*, vol. 14, no. 1, p. 3685, 2024. https://doi.org/10.1038/s41598-024-54078-w

[20] Y. Kong and H. Ding, "Tools, potential, and pitfalls of social media screening: Social profiling in the era of AI-assisted recruiting," *Journal of Business and Technical Communication*, vol. 38, no. 1, pp. 33–65, 2024. https://doi.org/10.1177/10506519231199478

[21] J. W. Creswell and J. D. Creswell, *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. Sage Publications, 2017.

[22] W. Bleidorn and C. J. Hopwood, "Using machine learning to advance personality assessment and theory," *Personality and Social Psychology Review*, vol. 23, no. 2, pp. 190–203, 2019. https://doi.org/10.1177/1088868318772990

[23] R. Baly *et al.*, "What was written vs. who read it: News media profiling using text analysis and social media context," *arXiv preprint* arXiv:2005.04518, 2020. https://doi.org/10.18653/v1/2020.acl-main.308

[24] S. M. Mohammad and P. D. Turney, "Crowdsourcing a word–emotion association lexicon," *Computational Intelligence*, vol. 29, no. 3, pp. 436–465, 2013. **https://doi.org/10.1111/j.1467-8640.2012.00460.x**

[25] F. Koto, A. Rahimi, J. H. Lau, and T. Baldwin, "IndoLEM and IndoBERT: A benchmark dataset and pre-trained language model for Indonesian NLP," *arXiv preprint* arXiv:2011.00677, 2020. https://doi.org/10.18653/v1/2020.coling-main.66

[26] A. Ghasempour, "Support vector regression to predict power consumption," *Electrical & Computer Engineering Technical Reports*, 2024.

[27] N. Reimers and I. Gurevych, "Sentence-BERT: Sentence embeddings using Siamese BERT-networks," in *Proc. EMNLP*, 2019, pp. 3982–3992. . https://doi.org/10.18653/v1/D19-1410

[28] P. L. Roth, P. Bobko, C. H. Van Iddekinge, and J. B. Thatcher, "Social media in employee-selection-related decisions: A research agenda for uncharted territory," *Journal of Management*, vol. 42, no. 1, pp. 269–298, 2016. https://doi.org/10.1177/0149206313503018

[29] M. Raghavan, S. Barocas, J. Kleinberg, and K. Levy, "Mitigating bias in algorithmic hiring: Evaluating claims and practices," in *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 2020, pp. 469–481. https://doi.org/10.1145/3351095.3372828

[30] J. Dastin, "Amazon scraps secret AI recruiting tool that showed bias against women," in *Ethics of Data and Analytics*. Auerbach Publications, 2022, pp. 296–299.

[31] B. H. Zhang, B. Lemoine, and M. Mitchell, "Mitigating unwanted biases with adversarial learning," in *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society*, 2018, pp. 335–340. https://doi.org/10.1145/3278721.3278779