



Article

Scalable Gender Profiling from Turkish Texts Using Deep Embeddings and Meta-Heuristic Feature Selection

Hakan Gunduz 

Software Engineering Department, Faculty of Engineering, Kocaeli University, Kocaeli 41001, Turkey;
hakan.gunduz@kocaeli.edu.tr

Abstract

Accurate gender identification from written text is critical for author profiling, recommendation systems, and demographic analytics in digital ecosystems. This study introduces a scalable framework for gender classification in Turkish, combining contextualized BERTurk and subword-aware FastText embeddings with three meta-heuristic feature selection algorithms: Genetic Algorithm (GA), Jaya and Artificial Rabbit Optimization (ARO). Evaluated on the IAG-TNKU corpus of 43,292 balanced Turkish news articles, the best-performing model—BERTurk+GA+LSTM—achieves 89.7% accuracy, while ARO reduces feature dimensionality by 90% with minimal performance loss. Beyond in-domain results, exploratory zero-shot and few-shot adaptation experiments on Turkish e-commerce product reviews demonstrate the framework's transferability: while zero-shot performance dropped to 59.8%, few-shot adaptation with only 200–400 labeled samples raised accuracy to 69.6–72.3%. These findings highlight both the limitations of training exclusively on news articles and the practical feasibility of adapting the framework to consumer-generated content with minimal supervision. In addition to technical outcomes, we critically examine ethical considerations in gender inference, including fairness, representation, and the binary nature of current datasets. This work contributes a reproducible and linguistically informed baseline for gender profiling in morphologically rich, low-resource languages, with demonstrated potential for adaptation across domains such as social media and e-commerce personalization.

Keywords: gender-aware user profiling; content personalization; BERT embeddings; meta-heuristic feature selection; LSTM classification; Turkish language processing



Academic Editors: Hyunchul Ahn,
Chienliang Lin, Wuke Zhang,
Sung-Wen Yu and Hanqing Xu

Received: 24 June 2025

Revised: 29 August 2025

Accepted: 9 September 2025

Published: 24 September 2025

Citation: Gunduz, H. Scalable Gender Profiling from Turkish Texts Using Deep Embeddings and Meta-Heuristic Feature Selection. *J. Theor. Appl. Electron. Commer. Res.* **2025**, *20*, 253. <https://doi.org/10.3390/jtaer20040253>

Copyright: © 2025 by the author. 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

Biological gender identification from written text has become an important task in natural language processing (NLP), with implications for author profiling, demographic analytics, and recommendation systems [1–4]. In digital platforms where user identity is partially disclosed or absent, inferring gender from linguistic patterns can improve downstream tasks such as content personalization, moderation, and segmentation [5,6].

Although many systems collect demographic information during registration, such data is often incomplete, outdated, or deliberately withheld. In such contexts, text-based gender inference plays a crucial role in addressing cold-start problems, where users generate content but lack a behavioral history. Although digital content spans modalities such as images, audio, and video, this study focuses exclusively on text as a representative and privacy-conscious medium. Compared to behavioral models based on user activity, textual

inference offers a lightweight and GDPR-compliant alternative, especially in systems that prioritize data minimization and ethical transparency.

Traditional approaches to gender classification have relied on shallow linguistic features such as bag-of-words or TF-IDF representations [3,7,8]. Although these models are computationally efficient, they often fail to capture semantic context, morphological structure, and long-range dependencies, particularly in morphologically rich languages like Turkish. Recent Transformer-based models such as BERT offer deep semantic embeddings [9,10], but their high-dimensional representations may introduce redundancy, noise, and overfitting risks without effective dimensionality reduction.

These challenges are compounded in low-resource agglutinative languages such as Turkish, where complex suffixation and lexical variation complicate syntactic parsing [1,11]. While deep embeddings have improved modeling, relatively few studies have explored dimensionally optimized representations for Turkish. Moreover, most previous work treats embedding spaces as static, ignoring the potential benefits of adaptive, model-aware feature selection, especially through meta-heuristic algorithms that balance efficiency with accuracy.

To address these gaps, we propose a hybrid gender classification framework that combines two complementary embeddings: BERTurk for contextual representation and FastText for subword-aware morphology, with three meta-heuristic feature selection methods: the Genetic Algorithm (GA), Jaya, and Artificial Rabbit Optimization (ARO). These embeddings are fed into two types of classifier: Long Short-Term Memory (LSTM) networks, which model sequential dependencies, and Gradient Boosting Machines (GBMs), which provide strong generalization and interpretability.

While our experiments are conducted on the publicly available IAG-TNKU corpus of 43,292 balanced Turkish news articles—chosen for its size, representativeness, and accessibility—the proposed framework is domain-agnostic and can be directly adapted to personalization scenarios in e-commerce, social media, and other digital ecosystems where textual user data is available but demographic attributes are missing.

The main contributions of this study are summarized below:

- We present a modular, scalable pipeline that integrates deep embeddings with meta-heuristic feature selection for gender classification in Turkish texts.
- We demonstrate that Artificial Rabbit Optimization (ARO) reduces the dimensionality of the features by 90% with minimal accuracy loss, allowing lightweight deployment.
- We show that the BERTurk+GA+LSTM configuration achieves state-of-the-art accuracy (89.7%) on the IAG-TNKU corpus.
- We perform comparative analysis between full-dimensional baselines and feature-optimized models, evaluating performance–efficiency trade-offs.
- We critically examine the ethical implications of text-based gender profiling, including fairness, generalizability, and binary gender assumptions.

The remainder of the paper is structured as follows: Section 2 reviews the prior literature on gender classification, embeddings, and feature optimization. Section 3 details the dataset, preprocessing, embedding models, feature selection algorithms, and classifiers. Section 4 presents the experimental setup and evaluation. Section 5 concludes the paper. Section 6 reflects on limitations, fairness, and directions for future work.

2. Theoretical Background and Related Work

2.1. Theoretical Foundations for Gender Classification

The proposed framework integrates contextualized (BERTurk) and subword-aware (FastText) embeddings with meta-heuristic feature selection methods to optimize gender

classification in Turkish news texts. This design is theoretically motivated by the linguistic characteristics of the Turkish, sociolinguistic and cognitive-linguistic theories of gendered language, the structure of modern embedding spaces, and the operational requirements for scalable classification models.

Turkish is an agglutinative language with rich morphology and highly productive suffixation. While Transformer-based models like BERTurk capture contextual dependencies, FastText represents each word as a bag of character n-grams, offering better generalization for rare or morphologically complex terms. However, both types of embedding produce dense and high-dimensional vectors, often exceeding 300 to 1024 features per sample. These high-dimensional spaces are susceptible to overfitting and can degrade computational efficiency.

Traditional dimensionality reduction techniques, such as Principal Component Analysis (PCA), are unsupervised and linear, projecting features into lower-dimensional latent spaces without direct regard for task-specific performance. These methods may fail to retain subtle yet critical gender-discriminative cues in complex NLP tasks. In contrast, meta-heuristic algorithms—such as GA, Jaya and ARO—are population-based global optimizers that perform task-specific searches to identify the most informative features. They are particularly effective in handling nonlinearity, sparsity, and high-dimensional spaces [12,13]. A comparative summary of PCA and meta-heuristic algorithms across various theoretical and practical dimensions is provided in Table 1, which highlights the motivations behind adopting meta-heuristic strategies in this study.

Table 1. Comparison of traditional and meta-heuristic feature selection techniques.

Aspect	Traditional Techniques (e.g., PCA)	Meta-Heuristic Algorithms (e.g., GA, Jaya, ARO)
Approach	Linear, unsupervised, projection-based	Population-based, global search, task-driven
Handling Nonlinearity	Limited	Strong (handles nonlinear, combinatorial spaces)
Feature Selection vs. Extraction	Feature extraction (creates new latent features)	Feature selection (chooses optimal subset from original features)
Task-Specific Optimization	Not tailored to classification objectives	Directly optimizes for classification performance
Scalability to High Dimensions	Can struggle with high-dimensional data	Designed for sparse, high-dimensional feature spaces
Computational Cost	Low	Higher (but yields better performance in complex settings)
Stability and Robustness	Moderate	Improved with variants (e.g., binary/informed meta-heuristics)

Meta-heuristic algorithms like GA, Jaya, and ARO have demonstrated strong performance in reducing feature space while preserving or improving classification accuracy, especially in domains involving high-dimensional embeddings and task-specific salience, such as text classification, sentiment analysis, and disease detection. Their flexibility and capacity to handle problem-specific evaluation functions make them particularly suitable for optimizing linguistic features extracted from embeddings. GA simulates evolutionary processes and has been applied in stylometric gender prediction [14]; Jaya is a parameter-free method with rapid convergence [15]; and ARO, inspired by rabbit population behavior, balances exploration and exploitation with low computational complexity.

In parallel, gender prediction from text is theoretically grounded in sociolinguistics and cognitive linguistics. Studies demonstrate that language use reflects and reinforces gender roles and social expectations [16,17]. Women tend to use more self-referent pronouns, social and sensory language, hedges, and indirect constructions, while men prefer impersonal forms, rarer vocabulary, and direct instrumental styles [18–20]. These patterns, while non-universal, are shaped by context, power dynamics, and social identity. They are linked to implicit and explicit attitudes [21] and can be mitigated through inclusive language practices [22].

While the present study adopts binary labels due to dataset constraints, it acknowledges the fluidity of gender expression and aims to advance fairness in NLP applications. By integrating linguistic theory, embedding representations, and optimization-driven feature selection, the framework aspires to balance accuracy, efficiency, and social responsibility in gender-aware computational modeling.

It is important to recognize that the theoretical motivations for our hybrid architecture, while grounded in linguistic and computational efficiency considerations, also intersect with broader ethical and fairness concerns in gender-aware NLP. The design choices we make in the selection, optimization, and deployment of models can influence not only classification performance but also how gendered language patterns are interpreted and implemented. Therefore, beyond methodological rigor, this work must be evaluated through the lens of fairness, transparency, and responsible AI, as discussed in the following subsection.

2.2. Fairness and Ethical Considerations

Gender identification from text is not only a technical challenge but is also a socially sensitive task, raising important questions about bias, fairness, and ethical responsibility in NLP. Recent research at the intersection of PAN (Plagiarism Analysis, Authorship Identification, and Near-Duplicate Detection), ACL (Association for Computational Linguistics), and gender-fair NLP have focused on identifying, measuring, and mitigating gender bias in language technologies. Shared tasks such as the Gendered Ambiguous Pronoun (GAP) challenge at the ACL Gender Bias in NLP Workshop have benchmarked systems' abilities to resolve pronouns fairly across genders, with top-performing models leveraging BERT-based architectures and ensemble methods to achieve near gender parity in coreference resolution [23].

Advances in author profiling, particularly in PAN competitions, demonstrate that deep learning models—especially those combining contextual embeddings with handcrafted features—can improve gender prediction accuracy and fairness over traditional statistical approaches [24]. However, studies also reveal that while debiasing techniques such as counterfactual data augmentation and sampling-based balancing can reduce bias according to certain fairness metrics, these improvements do not necessarily generalize across different definitions of fairness. This highlights the need for combined strategies that address both statistical fairness (equalized performance across groups) and causal fairness (removing bias from latent representations) [25].

Intersectional analysis further shows that existing debiasing strategies may fail to address compounded biases across multiple demographic attributes (e.g., gender and ethnicity) and in some cases can even amplify disparities in downstream tasks [26]. Comprehensive literature reviews emphasize the importance of acknowledging various forms of gender bias in NLP, critically assessing the strengths and weaknesses of mitigation strategies, and pursuing multifaceted solutions that integrate fairness objectives throughout the modeling pipeline [27].

In the present study, while the classification framework adopts binary gender labels due to dataset constraints, we acknowledge that gender identity is fluid and socially constructed. The methodology should therefore be interpreted as operating within the limits of available data rather than endorsing binary essentialism. By explicitly addressing these considerations, we aim to align our work with emerging standards for responsible AI, ensuring that technical performance improvements are balanced with ethical safeguards and fairness objectives.

2.3. Related Work

Biological gender identification from textual content has attracted substantial interest within author profiling, user analytics, and demographic inference tasks [6]. While early studies laid a solid foundation, recent advances in representation learning and feature optimization have reshaped the landscape of gender prediction systems.

2.3.1. Traditional Models and Shallow Representations

Initial efforts in gender classification relied mainly on shallow textual features such as bag-of-words, TF-IDF vectors, or stylometric indicators. In the Turkish context, Amasyali and Diri [1] reported an accuracy of 96.3% using Support Vector Machines (SVMs), while Yasdi and Diri [2] and Dogan and Diri [28] achieved similarly high results using SVM and classification based on n-gram. Comparable success was found in other languages; Alsmearat et al. [7] used stylometric features on Arabic news texts (80.4%), and Vijayakumar et al. [29] applied SVM and Naive Bayes to English Yelp reviews with 90.5% accuracy.

Despite high accuracy, these approaches often suffer from poor generalizability, limited semantic modeling, and sensitivity to domain shifts, especially in morphologically rich languages like Turkish.

2.3.2. Deep Learning and Embedding-Based Approaches

To overcome the limitations of shallow features, recent research has adopted deep learning methods and dense embeddings. For instance, Tufekci and Bektas [11] applied CNN and LSTM models to 13 Turkish datasets, achieving up to 96.7% in genre classification. Dalyan et al. [10] benchmarked deep models using FastText on Turkish texts and observed improved robustness over traditional pipelines. Bhagvati [9] and Abdallah et al. [30] used LSTM and SVM for gender classification in English, with accuracies above 82%.

Studies in Indonesian [31], Russian [3], and Arabic [32] have similarly demonstrated the effectiveness of neural architectures and embedding-based features across languages. Kucukyilmaz et al. [4] also introduced the notion of virtual gender to enrich binary classifiers in Turkish, reaching 85.4% accuracy.

2.3.3. User Metadata and Social Media Contexts

Complementary approaches have examined behavioral and contextual data for gender profiling. Onikoyi et al. [33] integrated metadata with linguistic features, improving accuracy by 10% on Twitter datasets. Angeles et al. [34] also used meta-attributes in conjunction with bag-of-words for social media classification. These studies highlight the value of multimodal signals but also underscore the challenge of relying on noisy or unavailable metadata in many real-world environments.

In addition to Turkish-language studies, extensive work has been carried out on English and European datasets within the PAN author profiling shared tasks and related Twitter-based studies. The PAN 2018 and 2019 competitions provided large multilingual Twitter corpora for gender identification, where top-performing systems combined stylometric and content-based features in ensemble frameworks, achieving up to 83.5% accuracy for English texts [35]. Earlier

text-only approaches, such as linear SVMs with n-gram and latent semantic analysis features, also performed strongly, reaching 82.2% accuracy in PAN 2018 [36].

More recent research has leveraged Transformer-based embeddings, particularly BERT, often combined with deep learning architectures. For example, Divya Jyothi [37] reported accuracies ranging from 78.5 to 87.5% for gender and personality prediction tasks using BERT and SimCSE embeddings with LSTM and CNN classifiers in the PAN2015 Twitter dataset. Other innovative preprocessing strategies, such as masking irrelevant terms while preserving text structure, improved accuracy to around 82% in English gender identification tasks [38].

Overall, these works highlight that deep contextual embeddings and hybrid modeling strategies consistently outperform traditional feature engineering approaches in English/European gender profiling. Integrating such insights strengthens the comparative scope of the present study by situating Turkish-language experiments within the broader international trend toward Transformer-based solutions in author profiling.

2.3.4. Recent Advances in Deep Learning for Gender Profiling in Multilingual and Low-Resource Contexts

A growing number of studies have explored the applicability of Transformer-based and hybrid neural architectures for author profiling in diverse linguistic settings. A recent study used BERT embeddings to capture contextual patterns for age and gender prediction and found that while TF-IDF combined with AdaBoost and KNN provided modest performance ($F1 = 0.58$), deeper embeddings have greater potential for semantic discrimination [39]. Another work applied a fine-tuned BERT model to the PAN 2018 dataset, achieving 79% accuracy, underscoring the efficacy of transfer learning in profiling tasks across social media domains [40]. CNN-BiLSTM attention hybrids have also been proposed, integrating clickstream text vectorized by CBOW with attention-enhanced architectures to improve the accuracy of commercial recommendations [41]. In Arabic, a novel dataset of 8000 articles from Modern Standard Arabic (MSA) was introduced along with an enhanced BERT model that incorporates gender-specific textual features. This approach yielded an accuracy 91% in MSA and 68.7% in dialectal text, demonstrating the value of semantically guided deep models across formal and informal genres [42].

Recent developments in unsupervised topic modeling have shown promise in interpreting embedding spaces. In particular, BERTopic [43] has emerged as a popular technique that combines Transformer-based embeddings with clustering and class-based TF-IDF to uncover coherent and diverse topics. Extensions of BERTopic have leveraged large language models (LLMs) to enhance topic quality and diversity in multilingual settings [44]. These methods not only improve semantic grouping but also help interpret the interpretability of embedding-based representations, an area relevant to gender profiling systems where feature transparency is crucial. Moreover, BERTopic has been applied in commercial domains such as IoT and e-commerce to detect latent trends, further illustrating its adaptability to domain-specific applications [45]. Although our framework focuses on supervised classification rather than unsupervised modeling, future extensions may benefit from integrating topic-level insights to explore gendered language patterns more transparently.

2.4. Research Gap and Positioning of This Study

While prior research demonstrates promising results using either deep embeddings or conventional classifiers, few studies have investigated the joint effect of contextual embeddings and meta-heuristic feature selection for gender classification. Moreover, most embedding-based methods operate on full-dimensional vectors, ignoring the computational overhead and overfitting risks associated with high-dimensional representations.

To the best of our knowledge, our study is the first to integrate BERTurk and FastText embeddings with three meta-heuristic algorithms—GA, Jaya and ARO—to optimize gender classification for Turkish. Unlike classical methods, our framework offers the following:

- a linguistically informed architecture tailored to Turkish morphology;
- algorithmic feature selection for real-time feasibility;
- cross-comparisons between feature selection techniques and classifiers.

By bridging deep contextual representation with scalable optimization, our work addresses both linguistic and computational challenges. This study contributes a reproducible and efficient framework for gender classification in the Turkish news content, with potential extensions to other low-resource and morphologically complex languages.

3. Methods

This study introduces a robust framework for gender-aware user profiling from text, designed to support personalization and customer segmentation in digital ecosystems. By leveraging advanced NLP techniques, the framework uncovers linguistic patterns associated with gender, an important demographic attribute in adaptive recommendation and content delivery. The approach integrates deep embedding models, meta-heuristic feature selection algorithms, and complementary machine/deep learning classifiers to extract, optimize, and evaluate discriminative features for predictive modeling. The methodology follows a six-stage pipeline: data acquisition, preprocessing, embedding-based feature extraction, optimization-driven feature selection, classification, and empirical validation. For clarity and reproducibility, a schematic flowchart of the pipeline is presented in Figure 1, illustrating the interaction between embeddings, feature selection, and classifiers.

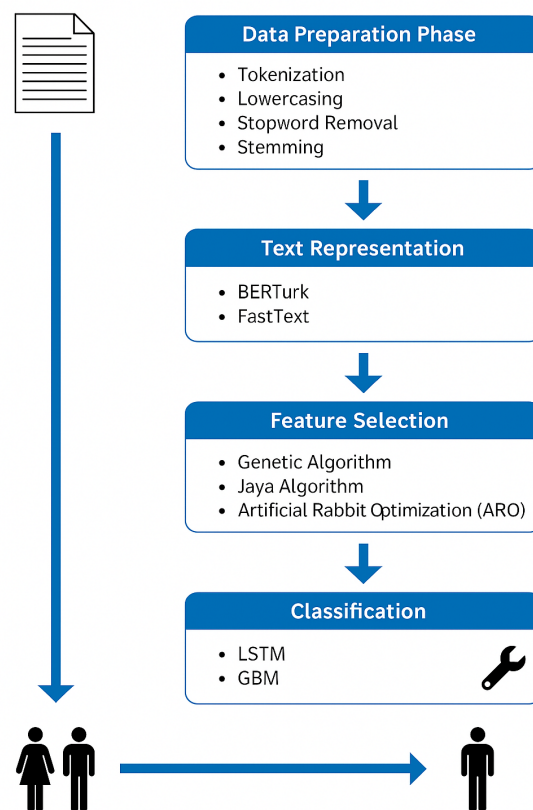


Figure 1. A visual flowchart of the proposed pipeline.

Building upon the theoretical foundations outlined in Section 2, this subsection details the practical motivations and empirical precedents guiding the selection of embedding models, feature selection algorithms, and classifiers in the proposed pipeline. The objective was to construct a modular and efficient system capable of handling high-dimensional, morphologically rich Turkish text while maintaining scalability.

Embedding Models. Two complementary strategies were employed:

- BERTurk, a Transformer-based model, captures deep contextual semantics and stylistic subtleties, making it well suited to gender classification tasks in Turkish.
- FastText represents words as bags of character n-grams, enabling robust handling of rare or morphologically complex tokens while remaining computationally efficient.

Feature Selection Algorithms. To mitigate redundancy and computational overhead in high-dimensional embeddings, three meta-heuristic algorithms were selected:

- Genetic Algorithm (GA), which leverages evolutionary principles to efficiently explore sparse feature spaces and avoid local minima;
- Jaya, a parameter-free method, which offers stable and deterministic convergence without the burden of hyperparameter tuning;
- Artificial Rabbit Optimization (ARO), which balances exploration and exploitation through biologically inspired strategies, demonstrating strong performance in recent high-dimensional optimization studies.

Classification Models. Two complementary classifiers were used to evaluate the discriminative capacity of optimized features:

- Long Short-Term Memory (LSTM) networks, which are capable of modeling sequential dependencies and stylistic cues across sentences.
- Gradient Boosting Machine (GBM), a robust ensemble method that is well suited for structured feature vectors, offering strong generalization.

This modular combination of embedding strategies, feature optimizers, and classifiers provides a balanced framework that advances both predictive accuracy and computational efficiency in Turkish gender classification tasks.

3.1. Dataset

This study employs the publicly available **IAG-TNKU dataset**, which was curated specifically for biological gender classification from Turkish news texts. The corpus consists of **43,212 news articles**, written by a total of **70 professional journalists—32 female and 38 male**. The dataset is perfectly balanced, containing 21,656 articles per gender, thus eliminating class imbalance and ensuring fair model training and evaluation.

Each article averages 429 words, with the entire corpus exceeding 22 million tokens. The document length distribution is right-skewed, with the majority of samples containing fewer than 1000 words, as illustrated in Figure 2. This makes it representative of medium-length editorial-style content in Turkish.

The dataset is freely available via the Mendeley Data repository at: <https://data.mendeley.com/datasets/8f93rjhgjk/1> (accessed on 20 April 2025).

Crucially, to avoid data leakage and prevent overfitting,

- The data set was first divided into 80% training and 20% test sets in a stratified manner based on gender labels. All feature selection procedures were applied only to the training set, and the resulting selected features were used to transform the test set for final evaluation.
- After the selection procedure, model training and performance evaluation in the data were performed using 10-fold stratified cross-validation (CV), ensuring a balanced gender distribution in all folds and a robust generalization analysis.

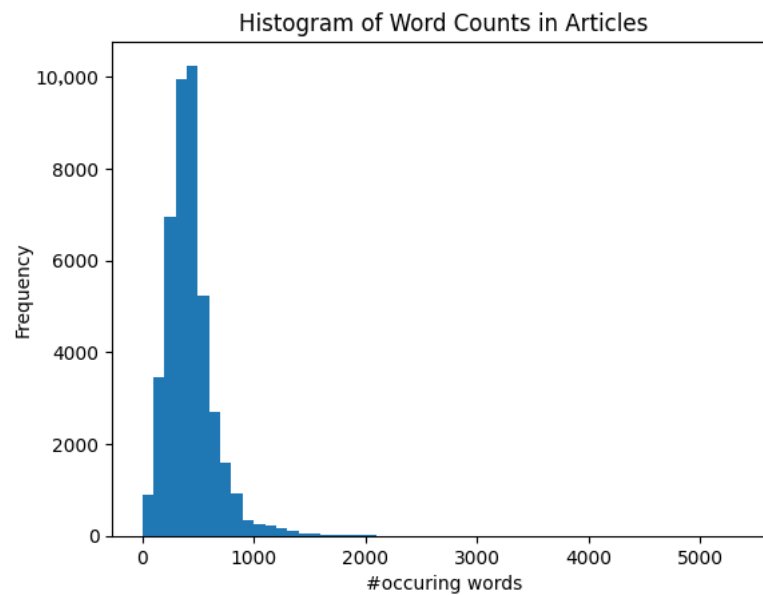


Figure 2. Histogram of word counts per article in the IAG-TNKU dataset.

3.2. Data Preprocessing

To ensure robust and standardized input for downstream NLP pipelines, a multi-stage preprocessing framework was applied to the raw dataset. This process aimed to normalize textual content, remove irrelevant noise, and ensure compatibility with deep embedding models and classifiers.

- **Unicode Normalization and Cleaning:** All articles were first normalized to UTF-8 format. Non-printable characters, broken characters, emojis, and HTML entities were removed.
- **Tokenization:** The corpus was tokenized at the word level using Turkish language rules to preserve morphological accuracy.
- **Lowercasing:** All tokens were converted to lowercase to reduce feature space and prevent case-induced sparsity.
- **Stopword Removal:** A custom Turkish stopwords list, based on the Zemberek NLP toolkit and domain-specific filtering, was used to remove non-discriminative words.
- **Punctuation and Symbol Filtering:** All punctuation marks, digits, and special characters were removed, except intra-word symbols relevant to named entities.
- **Stemming and Morphological Reduction:** The open-source Zemberek library was used for stemming and morphological analysis. This is critical for Turkish, a morphologically rich language, to reduce words to their root forms.
- **Length Normalization:** Articles were truncated or zero-padded to a maximum of **500 tokens** to support fixed-length input requirements of embedding models.
- **Embedding Preparation:** Two representations were generated:
 - The first are 768-dimensional BERTurk embeddings, using the dbmdz/bert-base-turkish-cased model.
 - The second are 300-dimensional FastText embeddings, computed as the average of all token vectors in the document.

3.3. Text Representation Using Deep Embeddings

Text representation plays a crucial role in converting raw textual data into meaningful numerical representations that can capture both linguistic and semantic nuances. In this study, we employ two advanced text representation models, BERT and FastText, which offer complementary strengths. BERT provides a deep contextual understanding

of the language, while FastText offers computational efficiency and robust handling of morphological variations.

3.3.1. Bidirectional Encoder Representations from Transformers (BERT)

BERT (Bidirectional Encoder Representations from Transformers) is a Transformer-based language model designed to capture bidirectional context by simultaneously attending to both left and right word sequences. It employs a masked language modeling strategy, enabling it to learn rich contextual dependencies within text [46].

The BERT architecture consists of stacked encoder layers composed of multi-head self-attention mechanisms and position-wise feed-forward networks. In addition, it incorporates segment and positional embeddings to differentiate sentence segments and preserve word order.

In this study, we utilized the pre-trained BERTurk model, specifically fine-tuned on Turkish corpora. Each article was tokenized into subword units using the BERT tokenizer, and the final outputs of the hidden layer were averaged to produce a 768-dimensional document vector. This vector was used for subsequent feature selection and classification tasks [47].

3.3.2. FastText

FastText, developed by Facebook AI Research, is a word embedding model that represents words as bags of character-level n-grams. This design enables it to effectively capture morphological patterns and handle out-of-vocabulary (OOV) words—an essential feature for agglutinative languages like Turkish [48].

In our implementation, we employed pre-trained 300-dimensional FastText embeddings trained on Turkish Wikipedia. Each article was converted into a single document vector by averaging the embeddings of all constituent words. Then these fixed-length vectors were forwarded to the feature selection and classification stages [49].

3.4. Feature Selection

In natural language processing tasks involving deep embeddings, models often encounter high-dimensional feature spaces. Although these representations are rich in semantic content, they can introduce redundancy, increase computational overhead, and increase the risk of overfitting. To address these challenges, we implement meta-heuristic feature selection techniques, which aim to identify the most informative subset of features that contribute significantly to classification performance.

Meta-heuristic algorithms are inspired by natural processes and are designed to efficiently explore large and complex search spaces. They offer flexible and robust solutions for optimizing feature selection by iteratively refining candidate subsets. In this study, we employ three such algorithms: GA, Jaya, and ARO. Each method explores the feature space differently, contributing to the discovery of optimal or near-optimal feature subsets.

3.4.1. Genetic Algorithm (GA)

The Genetic Algorithm (GA) [50,51] is a population-based optimization method modeled on the principles of natural evolution. It evolves a population of candidate solutions (feature subsets) through the application of genetic operators:

- **Selection:** High-fitness individuals are probabilistically chosen to contribute to the next generation.
- **Crossover:** Pairs of selected individuals are recombined to form offspring that inherit features from both parents.
- **Mutation:** Random alterations are introduced to maintain diversity and prevent premature convergence.

- **Elitism:** The best-performing solutions are preserved across generations to ensure consistent improvement.

The fitness of each solution is evaluated using classification performance metrics (e.g., F1 score or MCC). Over successive generations, GA converges toward feature subsets that enhance classification accuracy while minimizing dimensionality.

3.4.2. Jaya

Jaya [52] is a parameter-free optimization technique designed to move candidate solutions toward the best solution and away from the worst. Each candidate solution S_i^k at iteration k is updated using the following rule:

$$S_i^{k+1} = S_i^k + r_1 \cdot (S_{\text{best}} - |S_i^k|) - r_2 \cdot (S_{\text{worst}} - |S_i^k|) \quad (1)$$

Here, r_1 and r_2 are random variables uniformly sampled from $[0,1]$, and S_{best} and S_{worst} are the best and worst solutions in the population, respectively. In the context of feature selection, each solution represents a binary-encoded feature subset. The “best solution” refers to the subset achieving the highest classification performance (e.g., F1 score) on the current training fold, while the “worst solution” corresponds to the lowest-performing subset in the population. These definitions guide the algorithm update mechanism to more informative feature sets. This mechanism promotes the exploration of promising regions while avoiding suboptimal solutions. The process continues iteratively until a predefined stopping condition is met [53].

Jaya was selected over alternative meta-heuristics such as Particle Swarm Optimization (PSO) and Ant Colony Optimization (ACO) due to its parameter-free nature, which eliminates the need for tuning hyperparameters. This simplicity leads to stable and reproducible convergence in model parameter settings while maintaining competitive optimization performance.

3.4.3. Artificial Rabbit Optimization (ARO)

Artificial Rabbit Optimization (ARO) is a nature-inspired meta-heuristic proposed by Wang et al. [54], simulating the survival strategies of rabbits in the wild. The algorithm balances global exploration and local exploitation through two key behaviors—detour foraging and random hiding—regulated by an energy-based transition mechanism.

Detour Foraging (Exploration): When the energy factor is $A > 1$, rabbits forage away from their burrows to avoid detection. The position update rule is as follows:

$$X_i^{t+1} = X_i^t + R \cdot (X_i^t - X_r^t) + C \cdot \mathcal{N}(0,1) \quad (2)$$

where X_i^t is the position of the rabbit i , X_r^t is a randomly selected peer, R is a movement coefficient, and C introduces random perturbations.

Random Hiding (Exploitation): For $A \leq 1$, rabbits hide by randomly selecting one of several burrows:

$$b = X_i^t + H \cdot g \cdot X_i^t, \quad X_i^{t+1} = X_i^t + R \cdot (r_4 \cdot b - X_i^t) \quad (3)$$

where H is a linearly decreasing hiding coefficient, g is a binary mask, and r_4 is a random scalar.

Energy Shrink Mechanism: The energy factor A governs the balance of exploration and exploitation and is calculated as follows:

$$A = 2 \cdot \log(1/r) \cdot \theta, \quad \theta = 2 \cdot \left(1 - \frac{t}{t_{\text{max}}}\right) \quad (4)$$

Here, r is a uniform random number, t is the current iteration, and $t_{\max}$ is the maximum number of iterations.

The adaptive behavior of ARO and the stochastic exploration make it suitable for complex, high-dimensional optimization tasks such as feature selection. The application of ARO in this NLP context is particularly compelling because of the high-dimensional and sparse nature of textual embeddings. Detour foraging encourages global search across embedding dimensions, avoiding early convergence to suboptimal feature subsets. Meanwhile, random hiding facilitates focused refinement, enabling exploitation of promising feature combinations. The adaptive energy shrink mechanism dynamically adjusts the balance between exploration and exploitation based on iteration progress, aligning well with the evolving structure of feature importance in deep language representations.

The integration of GA, Jaya, and ARO into our feature selection pipeline allows for the discovery of compact and discriminative feature sets. These methods significantly contribute to reducing input dimensionality while preserving and often enhancing model accuracy and generalization.

3.5. Classification

We employed two state-of-the-art machine learning models well suited to handling high-dimensional and linguistically complex data: Long Short-Term Memory (LSTM) networks and Gradient Boosting Machines (GBMs). These models offer complementary advantages: LSTM excels at capturing sequential patterns in language, which are essential for modeling user behavior and stylistic cues, while GBM provides efficient handling of structured feature sets and strong performance on tabular representations. Their combined application enables robust and scalable analysis pipelines suitable for real-time recommendation, segmentation, and adaptive user experience systems on digital commerce platforms [55,56].

3.5.1. Long Short-Term Memory (LSTM)

LSTM networks are a specialized type of recurrent neural network (RNN) designed to model long-range dependencies in sequential data. They are particularly effective in text classification tasks, where understanding the order and context of words significantly impacts performance [55,56]. The core structure of an LSTM includes three types of gates—input, forget, and output—which collectively manage the flow of information through the network. These gates enable the model to selectively retain or discard information, allowing it to learn meaningful patterns across sequences.

In this study, LSTM networks were trained on sequential representations derived from BERT and FastText embeddings. We applied dropout regularization to mitigate overfitting and performed hyperparameter tuning via grid search to optimize key parameters such as the number of hidden units, learning rate, and batch size. The ability of LSTM to capture contextual flow makes it highly suitable for modeling the linguistic patterns associated with gender-specific writing styles.

3.5.2. Gradient Boosting Machine (GBM)

Gradient Boosting Machine (GBM) is a powerful ensemble learning method that constructs an additive model through a sequence of decision trees. Each new tree is trained to correct the residual errors made by the previous ensemble, thus refining the overall predictive performance [57,58]. GBM is particularly effective for structured data and has demonstrated robustness against overfitting when properly tuned. In our study, GBM was trained on feature vectors produced by applying dimensionality reduction (through feature selection) to the BERT and FastText embeddings. Hyperparameters such as the number of

estimators, the learning rate, and the maximum depth of the tree were optimized using a grid search.

Thanks to its iterative nature and strong generalization capability, GBM is well suited for capturing complex, nonlinear relationships between features and the target gender labels. Its efficiency and high accuracy make it a valuable baseline for comparison with deep learning approaches such as LSTM.

3.6. Performance Metrics

While accuracy is often suggested for evaluating performance, it fails to assess class discrimination. Alternatively, the F-measure serves as a valuable metric for validating class-based model assessments. Both the accuracy and the F-measure derivations are closely related to the confusion matrix (CM), which essentially indicates the count of correct and incorrect predictions for each class (Table 2). The metrics are derived using the values of true positive (tp), false positive (fp), false negative (fn), and true negative (tn) ([59]).

Table 2. Confusion matrix for two-class classification.

Actual/Predicted as	Positive	Negative
Positive	tp	fn
Negative	fp	tn

Accuracy is defined as the ratio of correct predictions to the total number of cases. However, when the number of false positives (fp) significantly exceeds false negatives (fn), it becomes necessary to employ the F-measure for performance evaluation. The F-measure computes the harmonic mean of precision and recall, thus considering both false positive and false negative cases to evaluate class separation. It is derived from the confusion matrix, as shown in Equations (5)–(7):

$$precision = \frac{tp}{tp + fp} \quad (5)$$

$$recall = \frac{tp}{tp + fn} \quad (6)$$

$$F\text{-Measure} = \frac{2 \times precision \times recall}{precision + recall}. \quad (7)$$

4. Experimental Results

This section presents the experimental procedures and evaluation results of the proposed gender classification framework. We detail the system configuration, hyperparameter optimization strategies, embedding configurations, and classification performance across different feature selection techniques and embedding types.

In the first stage, Turkish news articles and their corresponding author gender labels were collected from the IAG-TNKU dataset. To ensure linguistic consistency and reduce noise, the raw textual data underwent a standard preprocessing pipeline involving punctuation removal, tokenization, stopword elimination, and lemmatization.

In the second stage, the preprocessed texts were converted into dense vector representations using two pre-trained embedding models: BERTurk and FastText. BERTurk generates contextualized embeddings that capture semantic and syntactic dependencies, while FastText provides efficient subword-level embeddings that are well suited to morphologically rich languages such as Turkish.

The third stage involved dimensionality reduction through meta-heuristic feature selection techniques. To this end GA, Jaya and ARO were employed to identify compact

and highly discriminative feature subsets. This step significantly reduced model complexity while preserving predictive performance.

In the final stage, the selected features were used to train two types of classifiers: Long Short-Term Memory (LSTM) networks and Gradient Boosting Machines (GBMs). To ensure robust evaluation and minimize overfitting, all models were assessed using a 10-fold CV strategy. The entire experimental workflow was implemented in Python (v3.11.3), using TensorFlow (v2.4.1) and Huggingface Transformers (v4.5.1). All experiments were carried out on a machine running Ubuntu 20.04 LTS, equipped with an Intel Core i7-7700K CPU, 16 GB RAM, and an NVIDIA GTX 1070 GPU (8 GB). Environment management was handled using the Anaconda distribution.

4.1. Hyperparameter Optimization

To enhance the predictive performance of the LSTM and GBM classifiers, a systematic grid search strategy was employed to explore a predefined set of hyperparameter combinations. Model tuning was performed using 10-fold CV with grid search, where accuracy served as the primary evaluation metric. This approach ensured that the selected hyperparameters generalized well across data partitions and avoided overfitting.

Table 3 outlines the hyperparameter grid evaluated for the LSTM model, including variations in the number of memory units, dropout configurations, and activation functions. Table 4 presents the corresponding grid for the GBM classifier, featuring key parameters such as the number of estimators, the learning rate, and the depth of the tree.

In addition to classifier tuning, Table 5 provides the parameter settings adopted for the three meta-heuristic feature selection algorithms. These settings were selected based on the prior literature and validated through empirical testing on a held-out fold.

Table 3. Hyperparameter grid for the LSTM network.

Parameter	Values Tested
Number of memory cells	{30, 60, 120, 240}
Dropout rate	{0.2, 0.3, 0.4}
Recurrent dropout rate	{0.05, 0.1, 0.2}
Optimizer	Adam
Activation function	{ReLU, ELU}
Number of epochs	50
Batch size	32

Table 4. Hyperparameter grid for the GBM classifier.

Parameter	Values Tested
Number of estimators	{125, 250, 500, 1000}
Learning rate	{0.1, 0.01}
L2 regularization	{0.01, 0.001, 0.0001}
Maximum tree depth	{7, 9, 11, 13}

Table 5. Hyperparameter settings for meta-heuristic algorithms.

Algorithm	Population Size	Max Iterations	Other Parameters
GA	50	100	Crossover = 0.8, Mutation = 0.1
Jaya	50	100	None (parameter-free)
ARO	50	100	$R = 0.1$, $C = 0.2$, H linearly decayed

4.2. Feature Selection and Classification

In this study, we evaluated the impact of various feature selection strategies on the performance of two machine learning classifiers in the context of gender-aware user profiling. The feature selection process incorporated three meta-heuristic algorithms, each designed to optimize high-dimensional text representations for efficient deployment. These algorithms were applied to embeddings generated by two widely used pre-trained language models: FastText and BERT. By reducing the dimensionality of linguistic features without sacrificing predictive accuracy, our approach supports scalable, real-time demographic profiling on intelligent digital commerce platforms.

In the first experimental setting, FastText embeddings were used to convert Turkish news articles into dense vector representations. Each word in an article was mapped to a 300-dimensional vector using pre-trained FastText embeddings trained on Turkish Wikipedia. The documents were truncated or padded to 500 tokens, and the final representation of the document was the following:

- a 300-dimensional vector obtained by averaging the word embeddings (for GBM), or
- a 500×300 matrix input preserving token order (for LSTM).

In the second setting, contextualized embeddings were generated using the BERTurk model from the Huggingface Transformers library. The model was lightly fine-tuned in our dataset using a learning rate of 1×10^{-5} to preserve the pre-trained weights while adapting to domain-specific patterns. Tokenized input sequences (up to 500 tokens) were embedded into 768-dimensional vectors. Similar to the FastText configuration,

- GBM used the average of the 500 word-level vectors to create a single 768-dimensional input vector;
- LSTM operated on the full 2D input of 500×768 , capturing sequential dependencies.

To reduce redundancy and improve model generalization, the extracted embeddings were subjected to feature selection using the meta-heuristic algorithms. For each embedding type (BERT and FastText), a fixed-length document representation was first obtained by averaging word-level vectors—yielding a 768-dimensional vector for BERT and a 300-dimensional vector for FastText. These averaged vectors served as the input for the feature selection process, where the meta-heuristic algorithms searched for the most informative subset of features that maximized classification performance. Feature selection was performed independently for each type of embedding and classifier, allowing fine-grained optimization for both model architectures. Once the optimal feature subsets were selected, the resulting reduced vectors were reshaped and restructured into 2D matrices before being input to the LSTM model, thus preserving its ability to capture temporal dependencies. Meta-heuristic feature selection algorithms were assessed across 100 optimization epochs, with the Matthews Correlation Coefficient (MCC) and F1 score used as objective functions. The use of both MCC and F1 as optimization objectives ensures a balanced evaluation across class distributions, especially in scenarios where class imbalance may exist. The overall process was applied to both BERT and FastText embeddings, and the classification performance was evaluated using the LSTM and GBM models.

4.2.1. BERT Features

As summarized in Table 6, the **GA** achieved the highest accuracy with BERT embeddings. When optimized for MCC, GA selected 132 features and produced an accuracy of 89.7% for the LSTM model and 88.1% for GBM. Under F1 optimization, with 124 features, LSTM reached 88.8% and GBM 87.1%. **Jaya** also performed competitively, particularly under MCC optimization with 261 features, achieving 87.8% for LSTM and 87.2% for GBM.

When optimized for F1 (231 features), Jaya yielded slightly lower performance, with LSTM at 86.2% and GBM at 87.6%.

Table 6. Performance of meta-heuristic methods using BERT features.

Method	Optimization Metric	LSTM Accuracy $\pm$ Std	GBM Accuracy $\pm$ Std	#Features
GA	F1	0.888 $\pm$ 0.028	0.871 $\pm$ 0.036	124
GA	MCC	0.897 $\pm$ 0.033	0.881 $\pm$ 0.023	132
Jaya	F1	0.862 $\pm$ 0.023	0.876 $\pm$ 0.024	231
Jaya	MCC	0.878 $\pm$ 0.036	0.872 $\pm$ 0.018	261
ARO	F1	0.855 $\pm$ 0.025	0.861 $\pm$ 0.031	55
ARO	MCC	0.867 $\pm$ 0.033	0.868 $\pm$ 0.027	67

The **ARO** stood out for its efficiency, using only 67 features (MCC) and 55 features (F1) while maintaining strong performance. Specifically, ARO achieved 86.7% (LSTM) and 86.8% (GBM) under MCC optimization, and 85.5% (LSTM) and 86.1% (GBM) under F1 optimization. These results highlight the ability of ARO to significantly reduce dimensionality with minimal impact on classification accuracy.

4.2.2. FastText Features

Table 7 shows that similar trends were observed with FastText embeddings. The **GA** again delivered the best results, achieving 88.1% accuracy for LSTM and 87.1% for GBM using MCC optimization with 69 selected features. When optimized for F1 (53 features), GA reached 87.8% (LSTM) and 86.3% (GBM). The **Jaya**, using 119 features (MCC) and 104 features (F1), provided solid results, with LSTM accuracy of 86.7% and 85.5%, and GBM at 85.8% and 84.8%, respectively.

The **ARO** method once again demonstrated a balance between dimensionality reduction and performance. It achieved 85.8% (LSTM) and 85.1% (GBM) with 56 features (MCC), and 84.6% (LSTM) and 83.9% (GBM) with 44 features (F1).

Table 7. Performance of meta-heuristic methods using FastText features.

Method	Optimization Metric	LSTM Accuracy $\pm$ Std	GBM Accuracy $\pm$ Std	#Features
GA	F1	0.878 $\pm$ 0.035	0.863 $\pm$ 0.022	53
GA	MCC	0.881 $\pm$ 0.021	0.871 $\pm$ 0.033	69
Jaya	F1	0.855 $\pm$ 0.032	0.848 $\pm$ 0.024	104
Jaya	MCC	0.867 $\pm$ 0.028	0.858 $\pm$ 0.026	119
ARO	F1	0.846 $\pm$ 0.038	0.839 $\pm$ 0.034	44
ARO	MCC	0.858 $\pm$ 0.029	0.851 $\pm$ 0.024	56

5. Conclusions

This study presents a robust and modular framework for gender profiling in Turkish digital content, with the aim of enhancing personalization, segmentation, and intelligent user interaction in e-commerce ecosystems. By integrating deep contextual embeddings with meta-heuristic feature selection strategies, the proposed system delivers scalable and computationally efficient demographic inference.

Our experiments reveal that BERTurk embeddings consistently outperform FastText representations in capturing gender-discriminative linguistic signals. The best overall performance was achieved using a hybrid configuration of BERT embeddings, GA-based feature selection optimized with the MCC metric, and LSTM classification. This setup achieved an average accuracy of 89.7% in the IAG-TNKU corpus, establishing a new benchmark for gender classification in morphologically rich and under-resourced language settings.

To rigorously evaluate statistical robustness, we complemented 10-fold CV with Wilcoxon signed-rank tests, comparing each configuration against the LSTM+BERT+GA (MCC) baseline. Figure 3 presents the distribution of LSTM accuracies across configurations, while Table 8 summarizes the mean accuracies with 95% confidence intervals (CI) and corresponding p -values. The results herein confirm that the baseline configuration significantly outperforms most Jaya and ARO-based methods, with p -values frequently below 0.01. GA variants optimized with different metrics (F1, MCC) did not differ significantly, suggesting stability in GA performance between scoring functions.

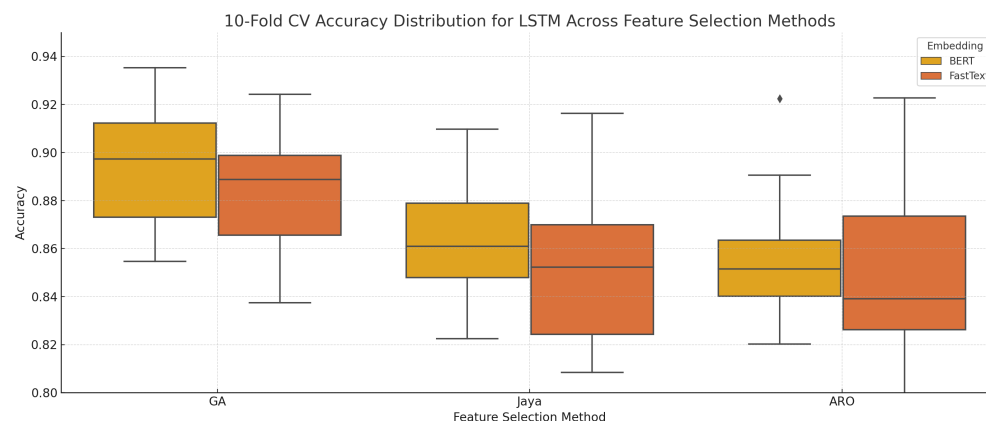


Figure 3. Representation of 10-fold CV accuracy distribution for LSTM across feature selection methods using BERT and FastText embeddings. GA consistently outperforms others with tighter variance.

Table 8. Representation of 10-fold LSTM accuracy with 95% CIs and Wilcoxon signed-rank p -values vs. the LSTM+BERT+GA+MCC baseline (two-sided).

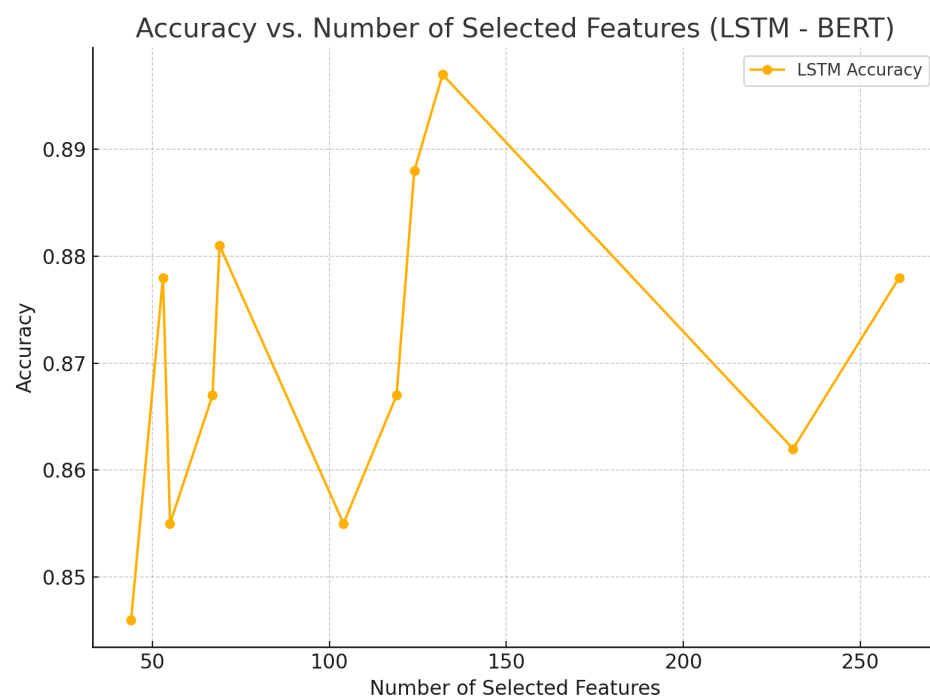
Model (BERT)	Accuracy (Mean $\pm$ 95% CI)	Wilcoxon p -Value
BERT+GA+F1	0.889 $\pm$ 0.017	0.9702
BERT+GA+MCC	0.895 $\pm$ 0.017	-
BERT+Jaya+F1	0.854 $\pm$ 0.013	0.0020
BERT+Jaya+MCC	0.879 $\pm$ 0.026	0.3223
BERT+ARO+F1	0.850 $\pm$ 0.011	0.0020
BERT+ARO+MCC	0.860 $\pm$ 0.021	0.0273
Model (FastText)	Accuracy (Mean $\pm$ 95% CI)	Wilcoxon p -Value
FastText+GA+F1	0.878 $\pm$ 0.022	0.1309
FastText+GA+MCC	0.888 $\pm$ 0.011	0.3223
FastText+Jaya+F1	0.851 $\pm$ 0.021	0.0039
FastText+Jaya+MCC	0.853 $\pm$ 0.023	0.0488
FastText+ARO+F1	0.834 $\pm$ 0.021	0.0020
FastText+ARO+MCC	0.862 $\pm$ 0.023	0.0371

We further assessed model generalizability by applying the same feature subsets to GBM, with statistical comparisons reported in Table 9. In nearly all cases, LSTM+BERT+GA (MCC) retained a statistically significant advantage over the GBM variants, reinforcing its suitability for sequential and contextual modeling in Turkish. Notably, even when differences were not statistically significant (e.g., BERT+GA+F1), LSTM maintained higher mean accuracy, underscoring its consistent edge.

Table 9. Representation of 10-fold GBM accuracy with 95% CIs and Wilcoxon signed-rank *p*-values vs. the LSTM+BERT+GA+MCC baseline (two-sided).

Model (BERT)	Accuracy (Mean $\pm$ 95% CI)	Wilcoxon <i>p</i> -Value
BERT+GA+F1	0.887 $\pm$ 0.019	0.5566
BERT+GA+MCC	0.863 $\pm$ 0.012	0.0273
BERT+Jaya+F1	0.871 $\pm$ 0.014	0.0840
BERT+Jaya+MCC	0.867 $\pm$ 0.015	0.0195
BERT+ARO+F1	0.853 $\pm$ 0.020	0.0195
BERT+ARO+MCC	0.873 $\pm$ 0.014	0.0488
Model (FastText)	Accuracy (Mean $\pm$ 95% CI)	Wilcoxon <i>p</i> -Value
FastText+GA+F1	0.863 $\pm$ 0.014	0.0137
FastText+GA+MCC	0.870 $\pm$ 0.032	0.0840
FastText+Jaya+F1	0.850 $\pm$ 0.013	0.0098
FastText+Jaya+MCC	0.854 $\pm$ 0.012	0.0039
FastText+ARO+F1	0.837 $\pm$ 0.021	0.0020
FastText+ARO+MCC	0.854 $\pm$ 0.021	0.0020

Accuracy trends over varying feature counts (Figure 4) further confirm the efficacy of feature selection, with optimal performance peaking around 140 features. Additionally, Figure 5 shows that GA not only achieves higher final accuracy but also converges faster than ARO when MCC is used as the optimization criterion.

**Figure 4.** Effect of number of selected features on LSTM accuracy using BERT embeddings. Accuracy peaks near 140 features.

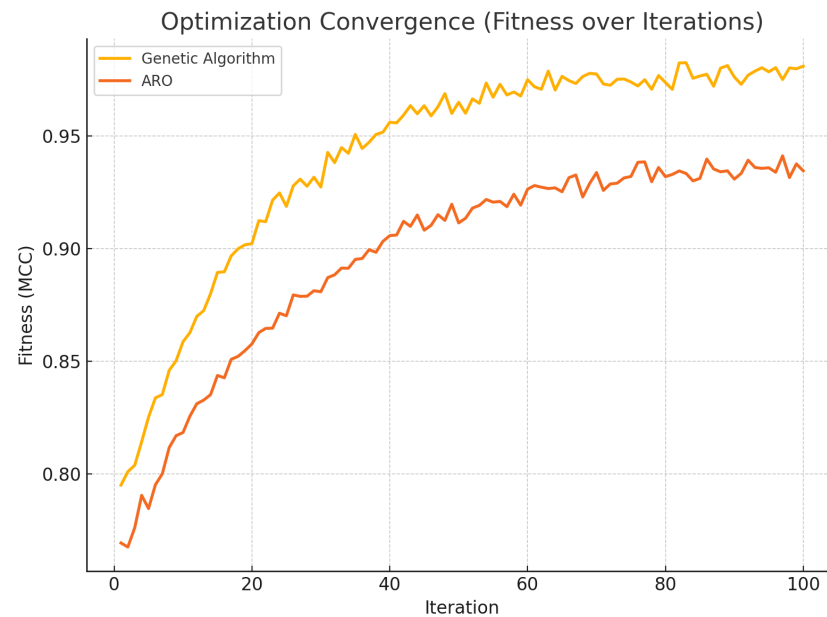


Figure 5. Convergence curves for GA and ARO over 100 iterations, using MCC as a fitness metric. GA shows faster convergence and higher final fitness.

Among all meta-heuristic methods, ARO stands out for its ability to reduce the feature set by up to 90% with only a modest drop in accuracy. Its biologically inspired search strategies proved effective for navigating sparse and non-convex embedding spaces. Although GA maintains the highest overall performance, ARO offers a compelling trade-off between compactness and predictive power, making it well suited to real-time and resource-constrained applications.

Our results confirm that LSTM-based models, when coupled with BERT embeddings and GA-driven feature selection, provide a statistically validated and high-performing solution for demographic inference in Turkish text. The integration of Wilcoxon signed-rank testing strengthens claims of robustness, demonstrating that observed gains are both numerically and statistically significant. The proposed framework, which is analytic in embedding and efficient, shows strong potential for gender-aware personalization across commercial and social platforms where fairness and scalability are critical.

6. Discussion

The findings of this study highlight both the strengths and boundaries of integrating deep embeddings with meta-heuristic feature selection for gender profiling in Turkish text. While the results confirm the effectiveness of our hybrid pipeline, they also reveal important considerations related to component-level contributions, cross-domain robustness, and the broader ethical landscape in which such models operate. The following subsections synthesize empirical insights from ablation studies, assess generalizability, position our contributions relative to prior work, and discuss limitations, operational constraints, and responsible use.

6.1. Ablation Studies and Component-Wise Evaluation

To assess the contributions of each pipeline component—including embeddings, feature selection algorithms, and dimensionality compression—we conducted a comprehensive ablation study. These experiments quantify the performance trade-offs involved and validate the design of our hybrid framework.

6.1.1. Embedding-Only Baselines (No Feature Selection)

We first evaluated the classification performance using the full-dimensional embeddings without any dimensionality reduction. These baseline configurations represent the upper bound of predictive accuracy but suffer from computational inefficiencies, making them less practical for real-time e-commerce applications.

- BERTurk (768D): LSTM achieved $91.1\% \pm 0.032$, GBM reached $89.3\% \pm 0.027$.
- FastText (300D): LSTM yielded $89.9\% \pm 0.028$, GBM achieved $88.3\% \pm 0.026$.

These results confirm the effectiveness of high-dimensional embeddings, but their latency and memory requirements pose challenges in resource-constrained or high-throughput scenarios.

6.1.2. Meta-Heuristic Feature Selection vs. Baseline

Next, we applied GA and ARO to reduce the feature space while retaining discriminative power. Both methods compressed the original embedding dimensions to between 44 and 132 features. This compression resulted in only a marginal drop in accuracy (typically 1–2%) compared to the embedding-only baselines (Figure 6).

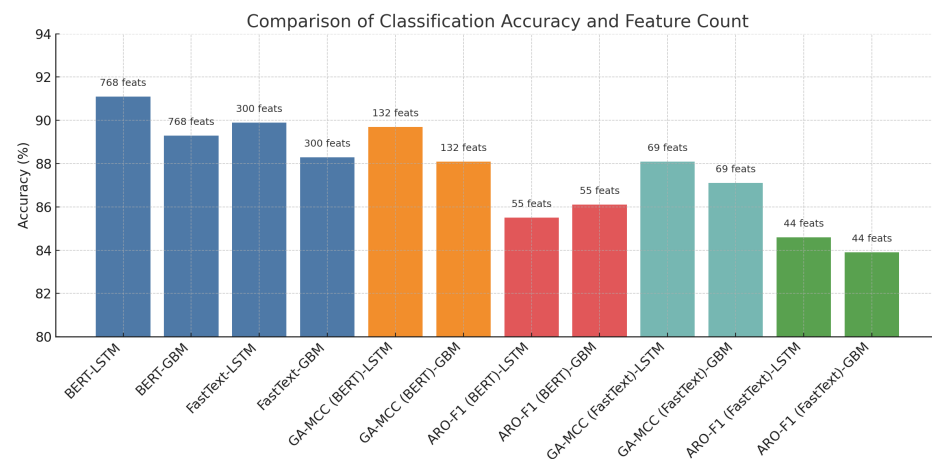


Figure 6. Comparison of classification accuracy between baseline models (using full-dimensional BERTurk and FastText embeddings) and models trained on feature-selected embeddings using GA and ARO. The number of features used in each configuration is indicated above each bar.

This trade-off highlights the practical utility of feature selection for real-world applications where speed, storage, and scalability are critical. Meta-heuristic methods enable the design of lightweight profiling systems suitable for recommender engines, customer segmentation, and adaptive interfaces.

6.1.3. Summary of Findings

The ablation analysis yielded the following key insights:

- Full-dimensional embeddings deliver high performance but are computationally expensive and less suitable for latency-sensitive systems.
- Meta-heuristic feature selection substantially reduces model size and inference time with minimal accuracy loss.
- ARO strikes a favorable balance between compression and predictive performance, validating its inclusion for lightweight profiling.

While these component-wise evaluations establish the computational and predictive trade-offs within our pipeline, real-world utility also depends on robustness beyond the controlled experimental domain. The next subsection examines how well our approach

might generalize to other genres, platforms, and linguistic contexts, as well as how its operational characteristics could be adapted to diverse application settings.

6.2. Zero-Shot and Few-Shot Applicability

Although the proposed framework demonstrates strong classification performance on the IAG-TNKU dataset, all samples originate from a single domain—Turkish news articles authored by professional journalists. This homogeneity introduces a limitation: the model's generalizability to other genres of user-generated or consumer-generated content (e.g., tweets, reviews, or support chats) remains untested in the primary evaluation.

To partially address this concern, we conducted exploratory zero-shot and few-shot experiments on a small set of Turkish e-commerce product reviews. This setting directly addresses concerns regarding cross-domain adaptability, since consumer-generated texts differ substantially from news articles in style, vocabulary, and noise levels (e.g., spelling errors and informal expressions).

In the zero-shot setup, the LSTM+BERT+GA (MCC) model trained exclusively on the IAG-TNKU news corpus was directly applied to the Turkish e-commerce reviews dataset (<https://www.kaggle.com/datasets/furkangozukara/turkish-product-reviews/data> (accessed on 24 August 2025)) [60], which consists of 1200 labeled samples (600 positive and 600 negative reviews). As reported in Table 10; without any adaptation, performance dropped to 58.9%, underscoring the challenge of domain transfer between professionally authored news articles and consumer-generated product reviews. However, when a limited number of labeled target-domain samples were introduced in a few-shot setting, accuracy improved noticeably to 69.6% with 200 labeled samples and 72.3% with 400 labeled samples. These preliminary results suggest that even modest supervision in the target domain can substantially mitigate domain shift, highlighting the feasibility of adapting our framework to practical personalization scenarios in e-commerce platforms.

Table 10. Exploratory zero-shot and few-shot results using the LSTM+BERT+GA+MCC configuration. Accuracy reported on Turkish e-commerce product review samples.

Evaluation Setting	Accuracy (%)
In-domain baseline (news corpus)	89.7
Zero-shot (e-commerce reviews, no adaptation)	58.9
Few-shot (200 labeled reviews)	69.6
Few-shot (400 labeled reviews)	72.3

These findings highlight two important points. First, the domain mismatch between professionally authored news and consumer-generated content is non-trivial. Second, few-shot learning provides a practical pathway for adapting our framework to real-world personalization scenarios (e.g., e-commerce) without requiring prohibitively large annotated datasets. Future work will extend these experiments to larger and more diverse consumer-generated corpora, as well as evaluate advanced domain adaptation techniques such as adversarial training or prompt-based LLM adaptation.

6.3. Significance in Relation to Prior Work

Although gender identification has been widely studied across different languages and platforms, our work introduces important methodological and practical contributions for AI-driven personalization in e-commerce. Unlike earlier approaches that rely on shallow lexical features and traditional classifiers such as SVM or Random Forest, our

pipeline integrates deep contextual embeddings with biologically inspired meta-heuristic optimization. This design enables scalable, interpretable, and accurate gender profiling.

For example, Tufekci et al. [59] applied LSTM to the same IAG-TNKU dataset using engineered features, achieving an accuracy of 88.51%. By leveraging BERT embeddings with GA-based feature selection, our model improved performance to 89.7%, reflecting the importance of contextual representation learning and semantic-aware feature optimization.

Compared to other state-of-the-art models—such as CNNs on Russian texts [8], Logistic Regression on Arabic tweets [32], or k-NN on English corpora [10]—our approach achieves superior or competitive accuracy. More importantly, it supports better scalability and interpretability for real-world integration in recommendation systems, user segmentation, and digital marketing pipelines. Table 11 provides a comparative summary.

Table 11. Performance comparison of our model with state-of-the-art gender identification studies.

Author	Language and Dataset Size	Classifier	Performance (%)
Sboev et al. [8]	Russian—1033 texts	CNN	88.0 (F1)
Hussein et al. [32]	Arabic—140,000 texts	LR	87.6 (Acc.)
Abdallah et al. [30]	Unspecified	SVM	82.81 (Acc.)
Kucukyilmaz et al. [4]	Turkish—835 texts	RF	85.40 (Acc.)
Dalyan et al. [10]	English—21,728 articles	k-NN	87.00 (F1)
Onikoyi et al. [33]	English—12,894 tweets	RF	70.00 (Acc.)
Tufekci et al. [59]	Turkish—43,292 articles	LSTM	88.51 (Acc.)
Our Study	Turkish—43,292 articles	BERT + GA + LSTM	89.7 (Acc.)

6.4. Limitations and Future Work

Despite promising results, the proposed framework has several important limitations. First, the study relies on a single dataset, the IAG-TNKU corpus of Turkish news articles. This corpus was selected because it is the largest publicly available and balanced Turkish dataset for author gender identification, containing more than 43,000 documents written by professional journalists. Its size, balance across genders, and accessibility make it uniquely suitable for benchmarking gender classification in Turkish. At the same time, we acknowledge that it represents a specific domain—editorial news writing—which differs linguistically and stylistically from consumer-generated content such as e-commerce reviews, tweets, or online forums. No large-scale, publicly available Turkish corpora of this kind currently exist, which limited our ability to directly test in-domain personalization scenarios. This constraint should therefore be viewed as a boundary condition of the present work rather than a general claim about all application settings.

Second, while our exploratory zero-shot and few-shot experiments on Turkish e-commerce reviews partially demonstrate adaptability, they also highlight that cross-domain transfer remains challenging. Performance dropped markedly in the zero-shot setting (58.9%) and improved only with few-shot fine-tuning (up to 72.3% with 400 reviews). These findings reinforce the importance of domain adaptation strategies and the need for more representative datasets covering consumer-generated text.

Third, the current framework focuses exclusively on textual signals, omitting potentially useful multimodal or contextual data such as metadata, behavioral features, or temporal activity. Incorporating these could further enhance robustness and personalization capabilities. Fourth, the meta-heuristic feature selection methods, while effective at reducing dimensionality, select latent embedding dimensions that lack direct linguistic interpretability. This reduces transparency, as the optimized feature subsets cannot be trivially mapped back to explicit lexical or syntactic patterns. Probing tasks and

post-hoc explainability tools (e.g., SHAP, LIME) should be explored in future work to strengthen interpretability.

Finally, the present study employs binary gender labels dictated by dataset availability, which risks oversimplifying gender as a social construct. Future extensions should incorporate non-binary and fluid gender categories, aligning with current research on gender-fair NLP and responsible AI.

6.5. Ethical Implications and Bias Mitigation

While our framework aims to support gender-aware personalization, automatic gender inference carries ethical risks, including bias amplification, stereotyping, and potential misuse in targeted advertising or content filtering. These risks are magnified in contexts with power asymmetries between service providers and users.

Per-class performance on the IAG-TNKU dataset was well balanced (male: 89.4% precision, 90.1% recall; female: 90.0% precision, 89.3% recall), suggesting minimal sensitivity disparity. However, balanced accuracy does not preclude the presence of bias in learned representations. Embeddings can encode social stereotypes, such as greater association of leadership-related terms with males, which could be inadvertently leveraged by the classifier.

To mitigate these risks, future iterations will incorporate the following:

- Post-hoc explainability to identify and audit potential stereotype-reinforcing features;
- Debaised embedding techniques (e.g., GN-GloVe, gender-neutral BERT) to reduce bias at the representation level;
- Fairness-aware optimization objectives that explicitly balance predictive performance and equity.

Ultimately, deploying such systems responsibly requires both technical safeguards and governance measures, ensuring transparency, user consent, and compliance with emerging AI ethics guidelines.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original data presented in the study are openly available in <https://data.mendeley.com/datasets/8f93rjhgjk/1> (accessed on 20 April 2025). To ensure full reproducibility of our experiments, all source code, preprocessing scripts, model configuration files, and statistical testing scripts have been made publicly accessible in the following GitHub repository: https://github.com/hakangunduz86/gender_identification_from_Turkish_texts (accessed on 24 August 2025).

Conflicts of Interest: The author declares no conflict of interest.

References

1. Amasyalı, M.F.; Diri, B. Automatic Turkish text categorization in terms of author, genre and gender. In Proceedings of the International Conference on Application of Natural Language to Information Systems, Klagenfurt, Austria, 31 May 2006; Springer: Berlin/Heidelberg, Germany, 2006; pp. 221–226.
2. Yasdi, M.; Diri, B. Author recognition by Abstract Feature Extraction. In Proceedings of the 2012 20th Signal Processing and Communications Applications Conference (SIU), Mugla, Turkey, 18–20 April 2012; IEEE: New York, NY, USA, 2012; pp. 1–4.
3. Sboev, A.; Litvinova, T.; Gudovskikh, D.; Rybka, R.; Moloshnikov, I. Machine learning models of text categorization by author gender using topic-independent features. *Procedia Comput. Sci.* **2016**, *101*, 135–142. [CrossRef]
4. Kucukyilmaz, T.; Deniz, A.; Kiziloz, H.E. Boosting gender identification using author preference. *Pattern Recognit. Lett.* **2020**, *140*, 245–251. [CrossRef]

5. Gülşen, E.; Gündüz, H.; Cataltepe, Z.; Serinol, L. Big data feature selection and projection for gender prediction based on user web behaviour. In Proceedings of the 2015 23rd Signal Processing and Communications Applications Conference (SIU), Malatya, Turkey, 16–19 May 2015; IEEE: New York, NY, USA, 2015; pp. 1545–1548.
6. Alves Gomes, M.; Meisen, T. A review on customer segmentation methods for personalized customer targeting in e-commerce use cases. *Inf. Syst. E-Bus. Manag.* **2023**, *21*, 527–570. [\[CrossRef\]](#)
7. Alsmearat, K.; Al-Ayyoub, M.; Al-Shalabi, R.; Kanaan, G. Author gender identification from Arabic text. *J. Inf. Secur. Appl.* **2017**, *35*, 85–95. [\[CrossRef\]](#)
8. Sboev, A.; Moloshnikov, I.; Gudovskikh, D.; Selivanov, A.; Rybka, R.; Litvinova, T. Automatic gender identification of author of Russian text by machine learning and neural net algorithms in case of gender deception. *Procedia Comput. Sci.* **2018**, *123*, 417–423. [\[CrossRef\]](#)
9. Bhagvati, C. Word representations for gender classification using deep learning. *Procedia Comput. Sci.* **2018**, *132*, 614–622. [\[CrossRef\]](#)
10. Dalyan, T.; Ayrar, H.; Özdemir, Ö. A comprehensive study of learning approaches for author gender identification. *Inf. Technol. Control* **2022**, *51*, 429–445. [\[CrossRef\]](#)
11. Tüfekci, P.; Bektaş, M. Author and genre identification of Turkish news texts using deep learning algorithms. *Sādhanā* **2022**, *47*, 194. [\[CrossRef\]](#)
12. Kurman, S.; Kisan, S. An in-depth and contrasting survey of meta-heuristic approaches with classical feature selection techniques specific to cervical cancer. *Knowl. Inf. Syst.* **2023**, *65*, 1881–1934. [\[CrossRef\]](#)
13. Kamal, R.; Amin, E.; Abdelminaam, D.; Ismail, R.M. A Comprehensive Survey on Meta-Heuristic Algorithms for Feature Selection in High-Dimensional Data: Challenges, Applications, and Future Directions. In Proceedings of the 2024 International Mobile, Intelligent, and Ubiquitous Computing Conference (MIUCC), Cairo, Egypt, 13–14 November 2024; pp. 84–90. [\[CrossRef\]](#)
14. Basyouni, A.; Abdelkader, H.; Elkilani, W.S.; Alharbi, A.; Xiao, Y.; Ali, A.H. A Suicidal Ideation Detection Framework on Social Media using Machine Learning and Genetic Algorithms. *IEEE Access* **2024**, *12*, 124816–124833. [\[CrossRef\]](#)
15. Zitar, R.A.; Al-betar, M.; Awadallah, M.A.; Doush, I.A.; Assaleh, K. An Intensive and Comprehensive Overview of JAYA Algorithm, its Versions and Applications. *Arch. Comput. Methods Eng.* **2021**, *29*, 763–792. [\[CrossRef\]](#)
16. Bigler, R.; Leaper, C. Gendered Language. *Policy Insights Behav. Brain Sci.* **2015**, *2*, 187–194. [\[CrossRef\]](#)
17. Menegatti, M.; Rubini, M. Gender Bias and Sexism in Language. *Oxf. Res. Encycl. Commun.* **2017**, *1*, 451–468. [\[CrossRef\]](#)
18. Brownlow, S.; Rosamond, J.A.; Parker, J. Gender-Linked Linguistic Behavior in Television Interviews. *Sex Roles* **2003**, *49*, 121–132. [\[CrossRef\]](#)
19. Mulac, A.; Bradac, J.; Gibbons, P. Empirical Support for the Gender-as-Culture Hypothesis: An Intercultural Analysis of Male/Female Language Differences. *Hum. Commun. Res.* **2001**, *27*, 121–152. [\[CrossRef\]](#)
20. Reid, S.A.; Keerie, N.; Palomares, N.A. Language, gender salience, and social influence. *J. Lang. Soc. Psychol.* **2003**, *22*, 210–233. [\[CrossRef\]](#)
21. Szczesny, S.; Moser, F.; Wood, W. Beyond sexist beliefs: How do people decide to use gender-inclusive language? *Personal. Soc. Psychol. Bull.* **2015**, *41*, 943–954. [\[CrossRef\]](#) [\[PubMed\]](#)
22. Irwandi, I.; Fajri, M.; Albert, A. Understanding gender and language use: A path towards effective interpersonal communication. *Mukadimah J. Pendidikan Sejarah Dan-Ilmu-Ilmu Sos.* **2024**. [\[CrossRef\]](#)
23. Webster, K.; Costa-jussà, M.R.; Hardmeier, C.; Radford, W. Gendered Ambiguous Pronoun (GAP) Shared Task at the Gender Bias in NLP Workshop 2019. In Proceedings of the First Workshop on Gender Bias in NLP, Florence, Italy, 2 August 2019. [\[CrossRef\]](#)
24. López-Santillán, R.; González, L.; Montes-y Gómez, M.; Lopez-Monroy, A.P. When attention is not enough to unveil a text's author profile: Enhancing a transformer with a wide branch. *Neural Comput. Appl.* **2023**, *35*, 9607–9626. [\[CrossRef\]](#)
25. Cyberek, H.; Ji, Y.; Evans, D. Addressing Both Statistical and Causal Gender Fairness in NLP Models. *arXiv* **2024**, arXiv:2404.00463. [\[CrossRef\]](#)
26. Lalor, J.; Yang, Y.; Smith, K.; Forsgren, N.; Abbasi, A. Benchmarking Intersectional Biases in NLP. In Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Seattle, DC, USA, 10–15 July 2022; pp. 3598–3609.
27. Sun, T.; Gaut, A.; Tang, S.; Huang, Y.; ElSherief, M.; Zhao, J.; Mirza, D.; Belding, E.; Chang, K.W.; Wang, W.Y. Mitigating Gender Bias in Natural Language Processing: Literature Review. *arXiv* **2019**, arXiv:1906.08976. [\[CrossRef\]](#)
28. Doğan, S.; Diri, B. Türkçe dokümanlar için N-gram tabanlı yeni bir sınıflandırma (Ng-ind): yazar, tür ve cinsiyet. *Türkiye Bilişim Vakfı Bilg. Bilim. Mühendisliği Derg.* **2010**, *3*, 11–19.
29. Vijayakumar, B.; Fuad, M.M.M. A new method to identify short-text authors using combinations of machine learning and natural language processing techniques. *Procedia Comput. Sci.* **2019**, *159*, 428–436. [\[CrossRef\]](#)
30. Abdallah, E.E.; Alzghoul, J.R.; Alzghool, M. Age and gender prediction in open domain text. *Procedia Comput. Sci.* **2020**, *170*, 563–570. [\[CrossRef\]](#)

31. Wongso, R.; Luwinda, F.A.; Trisnajaya, B.C.; Rusli, O. News article text classification in Indonesian language. *Procedia Comput. Sci.* **2017**, *116*, 137–143. [\[CrossRef\]](#)
32. Hussein, S.; Farouk, M.; Hemayed, E. Gender identification of egyptian dialect in twitter. *Egypt. Inform. J.* **2019**, *20*, 109–116. [\[CrossRef\]](#)
33. Onikoyi, B.; Nnamoko, N.; Korkontzelos, I. Gender prediction with descriptive textual data using a Machine Learning approach. *Nat. Lang. Process. J.* **2023**, *4*, 100018. [\[CrossRef\]](#)
34. Angeles, A.; Quintos, M.N.; Octaviano, M.; Raga, R. Text-based gender classification of twitter data using naive bayes and svm algorithm. In Proceedings of the TENCON 2021 IEEE Region 10 Conference (TENCON), Auckland, New Zealand, 7–10 December 2021; IEEE: New York, NY, USA, 2021; pp. 522–526.
35. Joo, Y.; Hwang, I. Profiling on Social Media: An Ensemble Learning Model using Various Features Notebook for PAN at CLEF 2019. In Proceedings of the CLEF (Working Notes), Lugano, Switzerland, 9 September 2019.
36. Daneshvar, S.; Inkpen, D. Gender Identification in Twitter using N-grams and LSA: Notebook for PAN at CLEF 2018. In Proceedings of the CLEF (Working Notes), Avignon, France, 10–14 September 2018.
37. Divya Jyothi, K. Author Profiling Approach: Predicting Personality Traits on Twitter Data using Combined BERT and SimCSE Embeddings. *Int. J. Res. Appl. Sci. Eng. Technol.* **2024**, *12*, 1140–1147. [\[CrossRef\]](#)
38. Jimenez-Villar, V.; Sánchez-Junquera, J.; Montes-y-Gómez, M.; Villaseñor-Pineda, L.; Ponzetto, S.P. Bots and Gender Profiling using Masking Techniques Notebook for PAN at CLEF 2019. In Proceedings of the CLEF (Working Notes), Lugano, Switzerland, 9 September 2019.
39. Sulochana, B.C.; Pragada, B.S.; Kiran, B.C.; Reddy, G.A.; Venugopalan, M. Author Identity Unveiled: Gender and Age Prediction from Textual Patterns using BERT. In Proceedings of the 2023 4th International Conference on Intelligent Technologies (CONIT), Hubballi, India, 21–23 June 2024. IEEE: New York, NY, USA, 2024; pp. 1–6.
40. Bsir, B.; Khoufi, N.; Zrigui, M. Prediction of author's profile basing on fine-tuning bert model. *Informatica* **2024**, *48*. [\[CrossRef\]](#)
41. Wang, Z.; Lu, X. Gender prediction model based on CNN-BiLSTM-attention hybrid. *Electron. Res. Arch.* **2025**, *33*. [\[CrossRef\]](#)
42. Himdi, H.; Shaalan, K. Advancing Author Gender Identification in Modern Standard Arabic with Innovative Deep Learning and Textual Feature Techniques. *Information* **2024**, *15*, 779. [\[CrossRef\]](#)
43. Grootendorst, M. BERTopic: Neural topic modeling with a class-based TF-IDF procedure. *arXiv* **2022**, arXiv:2203.05794.
44. Yang, C.; Kim, Y. Enhancing topic coherence and diversity in document embeddings using LLMs: A focus on BERTopic. *Expert Syst. Appl.* **2025**, *281*, 127517. [\[CrossRef\]](#)
45. Su, Y.S.; Wang, J.q.; Tu, S.H.; Liao, K.T.; Lin, C.L. Detecting latent topics and trends in IoT and e-commerce using BERTopic modeling. *Internet Things* **2025**, *32*, 101604. [\[CrossRef\]](#)
46. Devlin, J.; Chang, M.W.; Lee, K.; Toutanova, K. Bert: Pre-training of deep bidirectional transformers for language understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Minneapolis, MN, USA, 2–7 June 2019; (long and short papers); Volume 1, pp. 4171–4186.
47. Gündüz, H. Sentiment Analysis on Turkish Movie Reviews with Deep Bidirectional Encoder Representations from Transformers (Bert) And Support Vector Machines. *Kahramanmaraş Sutcu Imam Univ. J. Eng. Sci.* **2023**, *26*, 542–549.
48. Joulin, A.; Grave, E.; Bojanowski, P.; Douze, M.; Jégou, H.; Mikolov, T. Fasttext. zip: Compressing text classification models. *arXiv* **2016**, arXiv:1612.03651.
49. Sazan, S.A.; Miraz, M.H.; Rahman, A. Enhancing Depressive Post Detection in Bangla: A Comparative Study of TF-IDF, BERT and FastText Embeddings. *arXiv* **2024**, arXiv:2407.09187.
50. Osman, M.; He, J.; Zhu, N.; Mokbal, F.M.M. An ensemble learning framework for the detection of RPL attacks in IoT networks based on the genetic feature selection approach. *Ad Hoc Netw.* **2024**, *152*, 103331. [\[CrossRef\]](#)
51. Bohrer, J.d.S.; Dorn, M. Enhancing classification with hybrid feature selection: A multi-objective genetic algorithm for high-dimensional data. *Expert Syst. Appl.* **2024**, *255*, 124518. [\[CrossRef\]](#)
52. Rao, R.V.; Saroj, A. A self-adaptive multi-population based Jaya algorithm for engineering optimization. *Swarm Evol. Comput.* **2017**, *37*, 1–26.
53. Chaudhuri, A.; Sahu, T.P. Binary Jaya algorithm based on binary similarity measure for feature selection. *J. Ambient. Intell. Humaniz. Comput.* **2022**, *13*, 5627–5644. [\[CrossRef\]](#)
54. Wang, L.; Cao, Q.; Zhang, Z.; Mirjalili, S.; Zhao, W. Artificial rabbits optimization: A new bio-inspired meta-heuristic algorithm for solving engineering optimization problems. *Eng. Appl. Artif. Intell.* **2022**, *114*, 105082. [\[CrossRef\]](#)
55. Gunduz, H. Comparative analysis of BERT and FastText representations on crowdfunding campaign success prediction. *PeerJ Comput. Sci.* **2024**, *10*, e2316. [\[CrossRef\]](#)
56. Hu, Z.; Dychka, I.; Potapova, K.; Meliukh, V. Augmenting sentiment analysis prediction in binary text classification through advanced natural language processing models and classifiers. *Int. J. Inf. Technol. Comput. Sci.* **2024**, *16*, 16–31. [\[CrossRef\]](#)
57. Ghosal, S.; Jain, A. Depression and suicide risk detection on social media using fasttext embedding and xgboost classifier. *Procedia Comput. Sci.* **2023**, *218*, 1631–1639. [\[CrossRef\]](#)

58. Antonijevic, M.; Jovanovic, L.; Bacanin, N.; Zivkovic, M.; Kaljevic, J.; Zivkovic, T. Using bert with modified metaheuristic optimized xgboost for phishing email identification. In Proceedings of the International Conference on Artificial Intelligence and Smart Energy, Online, 22–23 March 2024. Springer: Berlin/Heidelberg, Germany, 2024; pp. 358–370.
59. Tüfekci, P.; Bektaş Kösesoy, M. Biological gender identification in Turkish news text using deep learning models. *Multimed. Tools Appl.* **2024**, *83*, 50669–50689. [[CrossRef](#)]
60. Gözükara, F.; Özel, S.A. Türkçe ve İngilizce Yorumların Duygu Analizinde Doküman Vektörü Hesaplama Yöntemleri için Bir Deneyisel İnceleme. *çUkurova üNiversitesi-Mühendislik-Mimar. Fakültesi Derg.* **2016**, *31*, 464–482. [[CrossRef](#)]

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.