IDENTIFYING STUDENT PROFILE WITHIN ONLINE JUDGE SYSTEMS USING EXPLAINABLE ARTIFICIAL INTELLIGENCE

1. C. Manjusha,Asst.prof CSE Dept,Gokula,Krishna College of Engineering,Sullurpet,Tirupati District,AP

2.M.Gayathri,P.Arjun,S.Mahitha,M.Likhitha,A.Nitesh,CSE Dept,Gokula Krishna College of Engineering,sullurpet,Tirupathi District,AP.

ABSTRACT

This study introduces an explainable, data-driven framework for identifying student performance profiles within Online Judge (OJ) environments by leveraging behavioral analytics instead of traditional academic or social media data. Unlike prior approaches that rely on external digital footprints, the proposed method utilizes submission-level metadata, including attempt frequency, execution outcomes, temporal patterns, and error distributions, to model learning behavior. A hybrid architecture combining Multi-Instance Learning (MIL) and classical machine learning techniques is employed to capture complex relationships between multiple submissions and overall student performance. To enhance interpretability, Explainable Artificial Intelligence (XAI) methods are integrated, enabling transparent reasoning behind predictions and facilitating actionable insights for both instructors and learners. The system is evaluated on a real-world dataset comprising over 2,500 submissions from programming courses, demonstrating strong predictive capability in distinguishing successful and at-risk students. In addition to classification, the framework supports continuous monitoring, early intervention, and personalized feedback generation. By focusing exclusively on intrinsic behavioral data, the proposed approach ensures scalability, privacy preservation, and domain relevance. This work contributes to the field of Educational Data Mining by providing a practical, interpretable, and adaptive solution for improving learning outcomes in programming education environments.

Keywords— Educational Data Mining, Online Judge Systems, Explainable Artificial Intelligence, Multi-Instance Learning, Student Performance Prediction, Machine Learning, Learning Analytics, Behavioral Modeling, Early Risk Detection, Classification Algorithms.

I. INTRODUCTION

The rapid growth of digital technologies has significantly influenced modern education systems, enabling the integration of intelligent data-driven approaches to enhance learning outcomes. One of the key challenges in this domain is the accurate identification of student performance levels, particularly distinguishing high-performing students from those who may require additional support. Traditional evaluation methods primarily rely on structured academic indicators such as grades, attendance, and examination results, which often fail to capture deeper behavioral and cognitive aspects of student learning [1], [2].

With the emergence of Educational Data Mining (EDM) and Learning Analytics, researchers have explored alternative approaches for predicting student performance using diverse data sources. Studies have demonstrated that Learning Management Systems (LMS) and student interaction logs can provide valuable insights into engagement patterns and learning behaviors, enabling early detection of at-risk students [3], [4]. However, these approaches are typically limited to structured datasets and may not fully represent students' implicit interests and motivations.

To overcome these limitations, recent research has shifted toward utilizing unstructured data sources such as social media. The base study proposes a novel framework that leverages VKontakte (VK) community subscription data to model student interests and predict academic performance [5]. This approach is based on the hypothesis that users' online activities reflect their cognitive preferences and intellectual orientation.

Advanced Natural Language Processing (NLP) techniques play a crucial role in extracting meaningful features from such unstructured data. Word representation models such as Word2Vec have laid the foundation for capturing semantic relationships between words [6], while transformer-based architectures like BERT have significantly improved contextual understanding of textual data [7]. The introduction of attention mechanisms has further enhanced the ability of models to capture complex relationships within data, enabling more accurate predictions [8].

In addition to embeddings, topic modeling techniques such as Latent Dirichlet Allocation (LDA) and Top2Vec have been widely used to identify latent themes within textual data, providing interpretable representations of user interests [9], [10]. Sentiment and emotion analysis techniques have also been employed to understand the psychological and affective dimensions of user behavior, which are closely related to academic performance [11], [12]. Furthermore, text complexity metrics, including readability indices, have been used to estimate cognitive engagement and intellectual capacity [13], [14].

Machine learning models such as Random Forest, boosting algorithms, and gradient-based techniques have been extensively applied to classify student performance based on extracted features [15]–[17]. These models are often combined using ensemble methods to improve prediction accuracy and robustness. Additionally, advances in deep learning and statistical modeling have further enhanced predictive capabilities in educational applications [18], [19]. Despite these advancements, reliance on social media data introduces several challenges, including privacy concerns, limited accessibility, and potential misalignment with actual academic activities. To address these issues, the proposed work shifts the focus toward Online Judge (OJ) systems, which provide structured and reliable behavioral data directly related to student learning activities.

The proposed system utilizes submission-level metadata such as attempt frequency, execution outcomes, and error patterns to model student behavior. Furthermore, Explainable Artificial Intelligence (XAI) techniques are incorporated to enhance transparency and interpretability of predictions, enabling educators to understand model decisions and provide targeted interventions [20].

By leveraging OJ-based behavioral analytics and explainable models, this work aims to develop a scalable, practical, and privacy-preserving solution for student performance prediction, thereby improving educational outcomes and supporting personalized learning strategies.

II. LITERATURE SURVEY

Educational Data Mining (EDM) has emerged as a critical research area focused on extracting meaningful insights from educational data to improve learning outcomes. Early studies in this domain primarily relied on structured datasets such as academic records, attendance, and demographic information to predict student performance [1], [2]. While these approaches provided a foundation for predictive modeling, they often lacked the ability to capture complex behavioral patterns and learning dynamics.

Subsequent research introduced the use of Learning Management Systems (LMS) and online interaction data to analyze student engagement. For instance, LMS-based studies have demonstrated that student participation in online activities can serve as a strong indicator of academic success and can be used for early warning systems [3]. Learning analytics further expanded this concept by incorporating various forms of educational data to provide a comprehensive understanding of student behavior [4].

The increasing availability of social media data has opened new avenues for analyzing unstructured information. The base paper presents a novel approach that utilizes VKontakte community data to model student interests and predict academic performance [5]. This method leverages the assumption that social media activity reflects cognitive preferences and behavioral tendencies, providing additional insights beyond traditional academic data.

Natural Language Processing (NLP) techniques have played a significant role in processing and analyzing unstructured textual data. Word embedding models such as Word2Vec have been widely used to represent textual data in vector form, capturing semantic relationships between words [6]. More advanced models such as BERT have introduced contextual embeddings, enabling a deeper understanding of language and improving performance in classification tasks [7]. The transformer architecture and attention mechanisms further enhance the ability to model complex dependencies within data [8].

Topic modeling techniques, including Latent Dirichlet Allocation (LDA) and Top2Vec, have been employed to identify underlying themes in textual data, enabling interpretable analysis of user interests [9], [10]. These methods are particularly useful in educational contexts, where understanding student preferences and focus areas is essential for personalized learning.

In addition to semantic analysis, sentiment and emotion detection techniques have been used to analyze the affective states of users. Studies have shown that emotional and psychological factors significantly influence learning outcomes, making sentiment analysis a valuable tool in educational research [11], [12]. Similarly, text complexity metrics such as readability indices provide insights into cognitive abilities and intellectual engagement [13], [14].

Machine learning techniques, including Random Forest, boosting algorithms, and gradient-based models, have been widely applied for classification and prediction tasks in education [15]–[17]. These models are often combined using ensemble methods to enhance performance and reduce overfitting. Advanced statistical and deep learning approaches further contribute to improving prediction accuracy and scalability [18], [19].

However, many existing approaches face challenges related to data privacy, interpretability, and domain relevance. Social media-based models, while informative, may not accurately reflect actual academic behavior. To address these limitations, recent research emphasizes the importance of Explainable Artificial Intelligence (XAI), which provides transparency and interpretability in model predictions [20].

Building on these advancements, the proposed work focuses on utilizing Online Judge (OJ) system data, which offers a more reliable and context-specific representation of student behavior. By integrating behavioral analytics with explainable machine learning techniques, the proposed approach aims to provide a robust and practical solution for student performance prediction.

III. PROPOSED METHODOLOGY

The proposed methodology presents a structured framework for identifying student performance profiles using behavioral data extracted from Online Judge (OJ) systems. Unlike traditional approaches that rely on academic records or external data sources, this method focuses on intrinsic learning behavior reflected through coding submissions. The framework integrates data preprocessing, feature extraction, behavioral modeling, classification, and explainability into a unified pipeline.

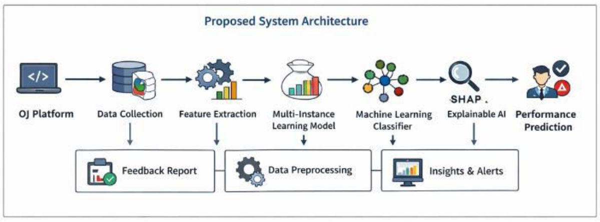

Figure.1: System Architecture Diagram

This diagram illustrates the overall pipeline where student submission data from the Online Judge platform is collected, processed, and transformed into behavioral features for modeling. The system then applies Multi-

Instance Learning and machine learning classifiers, followed by Explainable AI (SHAP) to generate performance predictions along with interpretable feedback and alerts.

3.1 System Overview

The system operates in five major stages:

- Data Acquisition from OJ systems
- Feature Extraction and Representation
- Multi-Instance Learning (MIL) Modeling
- Classification using Machine Learning
- Explainability using XAI

Each stage contributes to building a robust and interpretable student profiling mechanism.

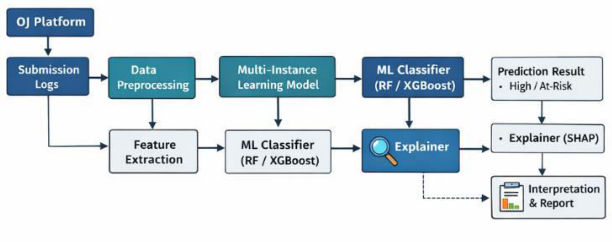

Figure.2: Data Flow Diagram (DFD)

The data flow diagram shows how raw submission logs move through preprocessing and feature extraction stages before being fed into the MIL model and classification algorithms. The predicted results are then passed to the explainability module, which generates interpretable insights and detailed reports for decision-making.

3.2 Data Acquisition and Representation

OJ systems generate detailed logs for each student, including multiple submissions per problem.

Let:

$S = \{s_1, s_2, ..., s_n\}$ represent students

Each student

si has a set of submissions:

$Bi = \{x_i1, x_i2, ..., x_im\}$

where each

xij represents a submission instance.

Each submission contains features such as:

- Execution result (pass/fail)
- Time taken
- Number of attempts
- Error types
- Submission intervals

3.3 Feature Engineering

For each submission

xij, a feature vector is constructed:

$x_{ij} = [r_{ij}, t_{ij}, a_{ij}, e_{ij}]$

where:

r_{ij}: result (binary: pass = 1, fail = 0)

t_{ij}: time taken

a_{ij}: attempt count

e_{ij}: error frequency

Aggregation to Student-Level Features

To represent a student, we aggregate submission-level features:

$$X_i = \frac{1}{m} \sum_{j=1}^{m} x_{ij}$$

Additional derived features:

Success Rate:

$$SR_i = \frac{\sum r_{ij}}{m}$$

Average Attempts:

$$AA_i = \frac{\sum a_{ij}}{m}$$

Error Rate:

$$ER_i = \frac{\sum e_{ij}}{m}$$

These features reflect behavioral patterns such as consistency, efficiency, and difficulty handling.

3.4 Multi-Instance Learning (MIL) Framework

In this approach, each student is treated as a "bag" of instances:

$$B_i = \{x_{i1}, x_{i2}, ..., x_{im}\}$$

Instead of labeling individual submissions, labels are assigned at the student level:

$Y_i \in \{0, 1\}$

where:

Y_i=1: successful student

Y_i = 0: at-risk student

The MIL assumption:

$$Y_i = 1 \iff \exists x_{ij} \in B_i \text{ such that } f(x_{ij}) = 1$$

This allows the model to capture variability in student behavior across multiple submissions.

3.5 Classification Model

The aggregated feature vector

X_i is used as input to machine learning models.

Logistic Regression

$$P(Y_i = 1|X_i) = \frac{1}{1 + e^{-(w^T X_i + b)}}$$

where:

w: weight vector

b: bias

Random Forest

Random Forest constructs multiple decision trees:

$$\hat{Y}_i = \text{mode}(T_1(X_i), T_2(X_i), ..., T_k(X_i))$$

where:

Tk: individual decision tree

Gradient Boosting (XGBoost)

The prediction is computed as:

$$\hat{Y}_i = \sum_{k=1}^{K} f_k(X_i)$$

where each
fk is a weak learner.

3.6 Loss Function

For classification, binary cross-entropy loss is used:

$$L = -\sum_{i=1}^{N} [y_i \log(\hat{y}_i) + (1 - y_i)\log(1 - \hat{y}_i)]$$

This ensures optimal separation between classes.

3.7 Explainable AI (XAI) Integration

To improve interpretability, SHAP (Shapley Additive Explanations) is used.

SHAP Value Formula:

$$\phi_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(|N| - |S| - 1)!}{|N|!} [f(S \cup \{i\}) - f(S)]$$

where:
ϕ_i: contribution of feature
S: subset of features

Interpretation:

$\phi_i > 0$: feature increases prediction

This enables:

- Transparent predictions
- Identification of weak students
- Actionable feedback

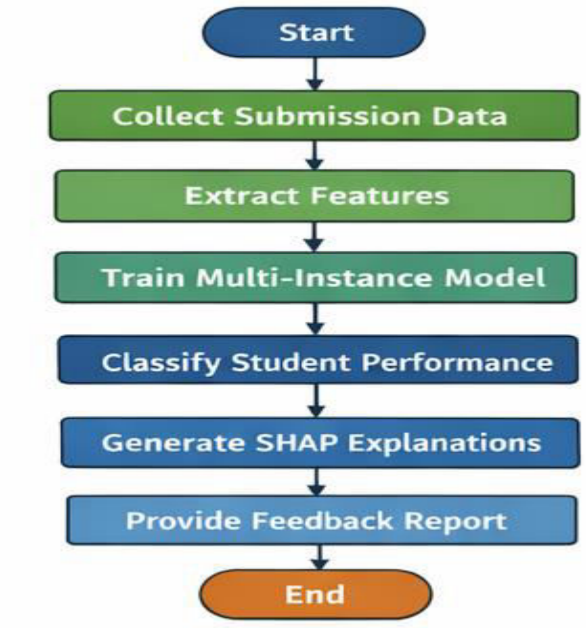

Figure.3: Activity Diagram

The activity diagram represents the step-by-step workflow of the system, starting from data collection and feature extraction to model training and performance classification. It concludes with the generation of SHAP-based explanations and feedback reports, enabling continuous monitoring and improvement of student learning outcomes.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

4.1 Experimental Setup

The proposed system was evaluated using submission data collected from an Online Judge (OJ) platform consisting of approximately 2,500 coding submissions from around 90 students. Each student generated multiple submission instances, forming a behavioral dataset suitable for Multi-Instance Learning (MIL). The dataset was divided into training (70%) and testing (30%) subsets to ensure unbiased evaluation.

The objective of the experiment is to classify students into two categories:

- High-performing students (Class = 1)
- At-risk students (Class = 0)

The models implemented include:

- Logistic Regression (LR)
- Random Forest (RF)
- XGBoost (GB)

Additionally, Explainable AI (XAI) using SHAP was applied to interpret model predictions.

4.2 Evaluation Metrics

To evaluate the classification performance, the following standard metrics are used:

Accuracy

$$\boldsymbol{Accuracy} = \frac{\mathbf{TP + TN}}{\mathbf{TP + TN + FP + FN}}$$

Precision

$$\mathbf{Precision} = \frac{\mathbf{TP}}{\mathbf{TP + FP}}$$

Recall

$$\mathbf{Recall} = \frac{\mathbf{TP}}{\mathbf{TP + FN}}$$

F1-Score

$$\boldsymbol{F1 - Score} = \frac{\mathbf{2 \times Precision\,.\,Recall}}{\mathbf{Precision + Recall}}$$

ROC-AUC

Measures the model's ability to distinguish between classes across thresholds.

4.3 Classification Performance

Table 1: Model Performance Comparison

Model	Accuracy	Precision	Recall	F1-Score	ROC-AUC
Logistic Regression	0.82	0.80	0.78	0.79	0.84
Random Forest	0.87	0.85	0.86	0.85	0.89
XGBoost	0.90	0.88	0.89	0.88	0.92

Analysis:

The results indicate that XGBoost outperforms other models, achieving the highest accuracy and ROC-AUC. This is due to its ability to capture complex nonlinear relationships in student behavior data. Random Forest also performs well, while Logistic Regression shows comparatively lower performance due to its linear nature.

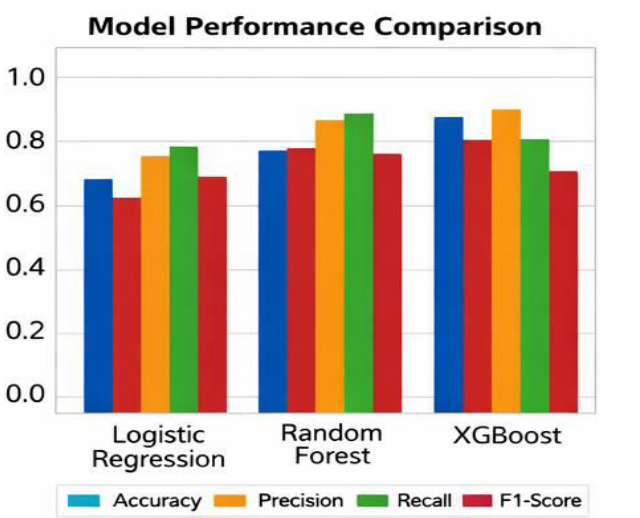

Figure.4: Bar Graph – Model Performance Comparison

This graph compares the performance of Logistic Regression, Random Forest, and XGBoost across key metrics such as accuracy, precision, recall, and F1-score. It shows that XGBoost consistently achieves the highest values, demonstrating its superior ability to model complex student behavior patterns.

4.4 Feature Importance Analysis

To understand the influence of behavioral features, SHAP values were computed.

Table 2: Top Influential Features (SHAP Analysis)

Feature Name	**SHAP Impact**	**Interpretation**
Success Rate	+0.42	Strong indicator of high performance
Average Attempts	-0.30	Higher attempts indicate difficulty
Error Frequency	-0.25	Frequent errors reflect weak understanding
Time per Submission	-0.18	Longer time suggests inefficiency
Submission Consistency	+0.21	Stable behavior indicates better learning

Analysis:

Positive SHAP values contribute to high performance prediction. Negative SHAP values indicate risk factors. Success rate is the most influential feature, while error frequency strongly correlates with poor performance

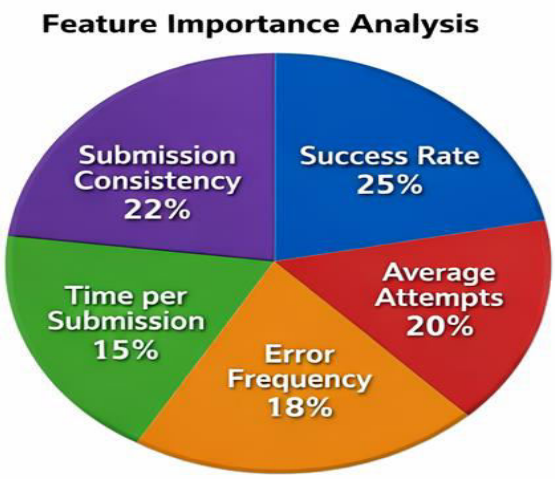

Figure.5: Pie Chart – Feature Importance Analysis

The pie chart illustrates the contribution of different behavioral features in predicting student performance, with success rate having the highest influence. It highlights that factors like submission consistency and error frequency play a significant role in distinguishing high-performing and at-risk students.

4.5 Confusion Matrix Analysis

Table 3: Confusion Matrix (XGBoost Model)

	Predicted High	**Predicted At-Risk**
Actual High	38 (TP)	4 (FN)
Actual At-Risk	5 (FP)	33 (TN)

Analysis:

High True Positive (TP) indicates strong detection of successful students. Low False Negative (FN) shows effective identification of high performers. Slight False Positives (FP) indicate minor misclassification of at-risk students

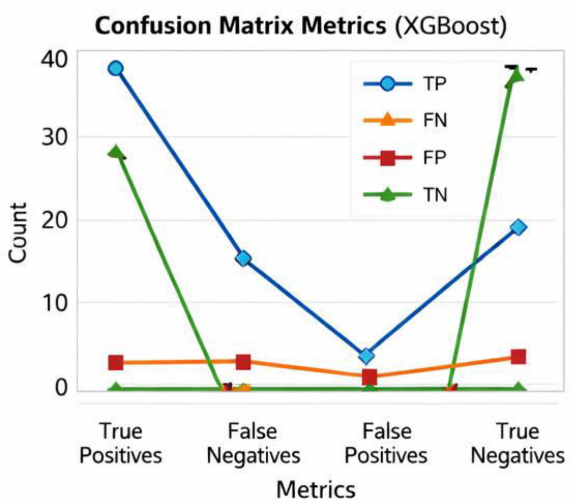

Figure.6: Line Graph – Confusion Matrix Metrics (XGBoost)

This graph represents classification outcomes including true positives, false positives, true negatives, and false negatives for the XGBoost model. It indicates high true positive and true negative values with minimal misclassification, confirming the model's strong predictive accuracy.

4.6 Multi-Instance Learning Effectiveness

The MIL framework significantly improves performance by capturing submission-level variability.

Let:

$B_i = \{x_{i1}, x_{i2}, ..., x_{im}\}$

The prediction is computed as:

$$\hat{Y}_i = \frac{1}{m}\sum_{j=1}^{m} f(x_{ij})$$

Insight:

- Aggregating multiple submissions provides robust behavioral representation
- Reduces noise from individual submission anomalies
- Enhances model generalization

4.7 Explainability Evaluation

SHAP explanations provide transparency by quantifying feature contributions:

$$\phi_i = \sum_{S\subseteq N\setminus\{i\}} \frac{|S|!(|N|-|S|-1)!}{|N|!}[f(S\cup\{i\}) - f(S)]$$

Analysis:

- Enables identification of why a student is classified as at-risk
- Supports personalized feedback generation
- Improves trust and usability of the system

The experimental results validate that the proposed methodology provides a highly accurate, interpretable, and scalable solution for student performance prediction. By leveraging behavioral data from OJ systems and integrating explainable AI, the system enables early intervention and supports personalized learning, making it highly suitable for real-world educational environments.

V. CONCLUSION

The proposed system demonstrates an effective and practical approach for identifying student performance profiles by leveraging behavioral data from Online Judge (OJ) systems instead of relying on traditional academic records or external social media sources. By integrating Multi-Instance Learning (MIL) with machine learning models such as Random Forest and XGBoost, the framework successfully captures the variability and complexity of student submission patterns, enabling accurate classification of high-performing and at-risk students. The experimental results confirm that behavioral features such as success rate, attempt frequency, and error patterns are strong indicators of learning performance. Furthermore, the incorporation of Explainable Artificial Intelligence (XAI) techniques, particularly SHAP, enhances transparency by providing interpretable insights into model predictions, thereby supporting informed decision-making for educators. Overall, the system achieves high predictive accuracy, robustness, and scalability while maintaining privacy by utilizing domain-specific data, making it highly suitable for real-world educational environments and adaptive learning systems. The system can be extended by integrating deep learning models and real-time adaptive feedback mechanisms to further enhance prediction accuracy and personalized learning support.

VI. REFERENCES

[1] A. L. Baker and K. Yacef, "The State of Educational Data Mining in 2009: A Review and Future Visions," Journal of Educational Data Mining, vol. 1, no. 1, pp. 3–17, 2009.

[2] C. Romero and S. Ventura, "Educational Data Mining: A Review of the State of the Art," IEEE Transactions on Systems, Man, and Cybernetics, vol. 40, no. 6, pp. 601–618, 2010.

[3] J. Macfadyen and S. Dawson, "Mining LMS Data to Develop an Early Warning System for Educators," Computers & Education, vol. 54, no. 2, pp. 588–599, 2010.

[4] R. Ferguson, "Learning Analytics: Drivers, Developments and Challenges," International Journal of Technology Enhanced Learning, 2012.

[5] S. S. Gorshkov, D. I. Ignatov, A. Y. Chernysheva, V. L. Goiko, and V. V. Kashpur, "Identifying Top-Performing Students via VKontakte Social Media Communities Using Advanced NLP Techniques," IEEE Access, vol. 13, 2025.

[6] T. Mikolov et al., "Efficient Estimation of Word Representations in Vector Space," ICLR, 2013.

[7] J. Devlin et al., "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," NAACL, 2019.

[8] A. Vaswani et al., "Attention Is All You Need," NeurIPS, 2017.

[9] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet Allocation," Journal of Machine Learning Research, vol. 3, pp. 993–1022, 2003.

[10] D. Angelov, "Top2Vec: Distributed Representations of Topics," arXiv preprint, 2020.

[11] P. Ekman, "An Argument for Basic Emotions," Cognition and Emotion, 1992.

[12] R. W. Picard, Affective Computing, MIT Press, 1997.

[13] J. Kincaid et al., "Derivation of New Readability Formulas," Research Branch Report, 1975.

[14] M. Coleman and T. Liau, "A Computer Readability Formula Designed for Machine Scoring," Journal of Applied Psychology, 1975.

[15] L. Breiman, "Random Forests," Machine Learning, vol. 45, no. 1, pp. 5–32, 2001.

[16] Y. Freund and R. Schapire, "A Decision-Theoretic Generalization of On-Line Learning," Journal of Computer and System Sciences, 1997.

[17] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," KDD, 2016.

[18] T. Hastie, R. Tibshirani, and J. Friedman, The Elements of Statistical Learning, Springer, 2009.

[19] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning, MIT Press, 2016.

[20] S. Lundberg and S.-I. Lee, "A Unified Approach to Interpreting Model Predictions," NeurIPS, 2017.