engineering proceedings

MDPI

Proceeding Paper

Sentiment Analysis of X Users in Digital Art: Comparison Between Algorithms †

Riana Magdalena Silitonga [1,*], Vivi Triyanti [1], Feliks Prasepta Sejahtera Surbakti [1], Devi Angrahini Anni Lembana [2], Valencia Catheryn Wilianto [1,*], Jennifer Angel Gala [1], Kayleen Gabreila [1] and Indah Munica Sari [1]

[1] Department of Industrial Engineering, Atma Jaya Catholic University of Indonesia, Sampora, Tangerang Selatan 15345, Indonesia; vivi.triyanti@atmajaya.ac.id (V.T.); feliks.prasepta@atmajaya.ac.id (F.P.S.S.); jennife.202304530016@student.atmajaya.ac.id (J.A.G.); vivkayleen.202304530003@student.atmajaya.ac.id (K.G.); indah.202304530018@student.atmajaya.ac.id (I.M.S.)

[2] Department of Management, Atma Jaya Catholic University of Indonesia, Jakarta 12930, Indonesia; devi.angrahini@atmajaya.ac.id

* Correspondence: riana.magdalena@atmajaya.ac.id (R.M.S.); valenci.202304530005@student.atmajaya.ac.id (V.C.W.)

† Presented at the 9th Eurasian Conference on Educational Innovation 2026 (ECEI 2026), Da Nang City, Vietnam, 30 January–2 February 2026.

Abstract

The rapid development of AI Technology has significantly influenced digital art and triggered widespread discussion on social media platforms, particularly X. The use of AI in generating visual artworks and digital content has elicited diverse public responses, ranging from support for technological innovation to concerns regarding originality and the role of human artists. In this study, a total of 1737 tweets were collected through a data crawling process using relevant keywords and processed using RapidMiner through preprocessing stages to analyze user sentiment on the X platform toward the application of AI in digital art. The data include data cleaning, text normalization, and tokenization, before being classified into positive and negative sentiments. Three classification algorithms, Naïve Bayes, support vector machine (SVM), and decision tree, were applied to compare sentiment distributions. The results show that the Naïve Bayes model classified 30.5% of tweets as positive and 69.5% as negative, while the SVM and Decision Tree models showed a stronger bias toward negative sentiment, with 93.3% and 88.8% negative classifications, respectively. These findings indicate that negative sentiment toward AI in digital art is more dominant among users.

Keywords: AI; digital art; decision tree; Naïve Bayes; support vector machine (SVM); sentiment analysis

Academic Editors: Teen-Hang Meen, Chun-Yen Chang and Cheng-Fu Yang

Published: 9 June 2026

1. Introduction

The rapid development of AI has reshaped artistic practices and the creative industries in profound ways. AI systems no longer function merely as technical tools; they have evolved into collaborative media capable of generating visual artworks, music, and various forms of digital content either autonomously or semi-autonomously [1]. This shift has sparked extensive debate regarding the position and role of human artists, particularly as computational models increasingly demonstrate the ability to imitate human creativity [2].

Previous research highlights that interactions between artists and algorithms introduce new dynamics to the creative process, where AI becomes not only a facilitator but also

a creative actor that contributes to shaping aesthetic outcomes [3]. Studies on artist–AI collaboration further emphasize that AI enhances visual exploration, accelerates creative workflows, and offers novel approaches unattainable through traditional techniques [4]. Nevertheless, the integration of AI into artistic production also raises critical concerns related to originality, plagiarism, artistic value, and the potential marginalization of human creative labor [5]. Many artists argue that AI-generated works blur the distinction between human expression and algorithmic output, giving rise to ethical and aesthetic tensions in contemporary art [6].

Similar phenomena are observed among fine arts students, who generally acknowledge AI's potential as an innovative tool while exhibiting ambivalence toward its implications for creative identity, art education, and future career prospects. While students regard AI as a means to broaden visual experimentation, others fear that reliance on such technology may diminish fundamental artistic skills essential to human creativity [7].

From the perspective of digital and design creators, the growing use of AI is often met with skepticism rather than dogmatism. Although many acknowledge its usefulness in speeding up production and offering visual variations, there is concern that these conveniences come at a creative cost. Designers increasingly worry that AI-driven outputs encourage standardized aesthetics and quick solutions, which can limit originality and reduce the depth of creative exploration [8]. When design decisions rely heavily on automated systems, the role of personal interpretation and intentional choice risks being weakened.

Beyond artistic communities, public discourse on social media has become increasingly active in responding to the rise of AI-generated art. Platforms such as X serve as dynamic spaces where users express diverse opinions, concerns, and support related to the use of AI in digital art [9]. Prior studies on public sentiment in Twitter-based discussions indicate that social media plays a significant role in capturing collective perceptions and emotional reactions toward emerging technologies, making sentiment analysis a relevant method for understanding public opinion within the context of technological change [10].

Given the growing intensity of online debates and the polarized attitudes toward AI's role in digital art, examining user sentiment on X has become increasingly important [11]. The results of this study provide a reference to understand public perceptions of AI within the creative domain, contributing to the methodological development of machine-learning-based sentiment analysis, which holds broader applicability across various social and technological contexts.

2. Literature Review

2.1. Text Mining

Text mining is the process of analyzing text-based data to extract meaningful information from unstructured sources [12]. This technique is becoming increasingly important because most digital data is in the form of text, such as social media posts, reviews, articles, and online conversations. The text mining process involves several stages, from data acquisition, text cleaning, and transformation, to feature extraction using specific methods [13]. With text mining, text data that was originally random and unstructured can be converted into relevant information for further analysis.

Text mining is used to uncover hidden patterns, trends, and relationships within large volumes of textual data that would be difficult to analyze manually. By applying techniques such as tokenization, stemming or lemmatization, term frequency analysis, and natural language processing (NLP), text mining supports tasks, including sentiment analysis, topic modeling, and text classification [14]. These outputs can then be used to support decision-making, improve services or products, and gain deeper insights

into public opinions and behaviors. As a result, text mining not only enhances the efficiency of data analysis but also adds strategic value by transforming raw textual data into actionable knowledge.

In scientific research, text mining is widely used to detect patterns, trends, or opinions from a group of users. When applied to social media, text mining allows researchers to understand the dynamics of public discussions on a particular issue, including public perceptions of technological developments such as AI [15]. This technique is also a key foundation in machine learning-based analysis, as it provides data in a form that is ready to be processed for classification or prediction.

Furthermore, text mining is also used to process large-scale and real-time data efficiently, especially in fast-evolving platforms like social media, where information changes rapidly. By combining text mining with machine learning algorithms, researchers can monitor shifts in public sentiment, identify emerging topics, and detect anomalies or misinformation related to AI and other technologies [16]. This integration enhances the reliability and depth of analysis, enabling more accurate interpretations of societal responses and supporting evidence-based conclusions in scientific studies.

2.2. Sentiment Analysis

Sentiment analysis is conducted to identify emotional or opinion trends in a text, which are generally classified into positive, negative, or neutral categories. This technique is applied in marketing, politics, education, and technology because it provides a quantitative overview of public attitudes toward specific topics or issues [17]. By analyzing large volumes of textual data, sentiment analysis helps researchers and organizations understand how people emotionally respond to products, policies, or technological developments.

In research based on short texts such as tweets, sentiment analysis presents particular challenges. Users often express their opinions briefly and informally, frequently using slang, abbreviations, emojis, or metaphorical language. These characteristics can make it difficult for analytical models to accurately interpret the true sentiment, especially when sarcasm or implicit meaning is involved. As a result, careful preprocessing and model selection are crucial to improve analysis accuracy [18].

Sentiment analysis methods adopt two approaches: lexicon-based and machine learning-based methods. The lexicon-based approach relies on predefined dictionaries of words with associated sentiment values, while the machine learning approach learns sentiment patterns directly from labeled data [19]. Currently, machine learning methods are favorable because they are more flexible and capable of capturing complex linguistic patterns within textual data.

In digital art, sentiment analysis using AI plays an important role in assessing public responses to the use of AI in creative processes. It helps determine whether public opinion tends to support, reject, or critically evaluate the role of AI in the creative world. The insights generated from this analysis are valuable for understanding societal acceptance and concerns, which may influence future industry development, policy decisions, and artistic practices [20].

2.3. X Platform as Data Source

X, formerly known as Twitter, is a text-based social media platform that provides an open space for public discussion on a wide range of issues. Its real-time and publicly accessible nature allows users to instantly share opinions, reactions, and information, making the platform highly dynamic. These characteristics make X particularly suitable for research that focuses on public discourse and opinion formation [21].

One of the main advantages of X is its concise text format, which encourages users to express ideas briefly and directly. This creates a large volume of short text data that is rich in opinions and emotional expressions. For researchers, this format is valuable for analyzing how public sentiment emerges and evolves in response to specific events or topics [22]. In addition, features such as hashtags, mentions, and retweets play an important role in shaping conversations on X. Hashtags help categorize discussions and identify trending topics, while mentions and retweets allow interactions and information diffusion among users. These features enable researchers to trace conversation flows, detect dominant themes, and map relationships between topics and users [23].

Furthermore, X has a diverse user base that represents a wide range of social, cultural, and professional backgrounds. On technology-related topics such as AI, users often engage in interactive exchanges that bring together different perspectives. Using X as a data source, therefore, allows researchers to capture the dynamics of public opinion more accurately, particularly in debates on the development of AI in digital art [24].

2.4. Crawling Data on Social Media

Data crawling is used to collect data from the internet or specific digital platforms through scripts or specialized software [25]. In social media–based public opinion research, data crawling functions as a primary method for gathering large volumes of textual data, such as posts, comments, and user interactions [26]. This automated approach enables researchers to efficiently obtain datasets that would be impractical to collect manually, especially when dealing with real-time or large-scale data sources.

Through data crawling, specific parameters such as keywords, hashtags, user accounts, or topics of interest can be defined [27]. This targeted collection process ensures that the retrieved data is relevant to the research objectives and appropriate for further analysis. As a result, crawling plays a crucial role in preparing raw data that can later be processed through text mining, sentiment analysis, or machine learning techniques.

In research on Platform X, data crawling is essential because the content is dynamic and continuously updated. Researchers apply filters such as time range, language, and geographic indicators to capture discussions within a specific context. These settings help maintain the relevance and consistency of the dataset, enabling more accurate analysis of public opinion over time.

Despite its advantages, data crawling also presents challenges. Limitations imposed by platform APIs, frequent changes in platform structures, and the presence of noise or irrelevant data can reduce data quality [28]. Additional preprocessing and filtering steps are performed to improve reliability. Nevertheless, data crawling remains a fundamental stage in text mining and sentiment analysis, as it provides the essential foundation for subsequent analytical process.

2.5. Decision Tree (DT)

A DT is a classification algorithm that divides data into a branching structure resembling a DT [29]. Each node in the tree represents a condition or feature, while each branch shows the result of testing that condition. This model is considered easy to understand because it reflects human decision-making logic, making it appropriate as a baseline in machine learning research. In text analysis, decision trees distinguish linguistic patterns that indicate positive or negative sentiment.

Despite its simplicity, DTs have weaknesses such as a tendency to overfit, especially with high-dimensional data such as text. However, when combined with optimization techniques such as pruning, their performance can improve significantly. In social media research, DTs are used to understand which features most influence sentiment classifica-

tion [30]. This makes DTs relevant as a comparison to other more complex algorithms, such as Naïve Bayes.

In addition, DTs offer a level of interpretability that often lacks in more complex machine learning models. DTs enable the observation of how specific features contribute to classification outcomes, which is valuable in text-based studies where understanding the reasoning behind a prediction is as important as the prediction itself. This transparency allows analysts to identify key words, phrases, or linguistic structures that strongly affect sentiment, thereby supporting more meaningful qualitative interpretations of quantitative results [31].

Furthermore, DTs are relatively efficient in terms of computation and do not require extensive parameter tuning compared to advanced models. This makes them practical for exploratory analysis or for studies with limited computational resources. Although their predictive accuracy may be lower than that of more sophisticated algorithms, DTs remain useful as a benchmark model, providing a clear point of reference when evaluating the performance and added complexity of alternative approaches [32].

2.6. Support Vector Machine (SVM)

SVM is a popular classification algorithm in text processing due to its ability to handle high-dimensional data and non-linear classification [33]. SVM classifies data by identifying the optimal hyperplane that maximizes the separation between classes. In sentiment analysis, SVM is effective because it can detect complex patterns in text representations such as TF-IDF vectors or word embeddings. Its strong generalization ability has made SVM one of the most widely used algorithms in natural language processing research. SVM is also robust against overfitting, particularly with sparse and high-dimensional text data [34]. By applying kernel functions, which are linear, polynomial, or radial basis functions, SVM transforms input data into higher-dimensional spaces where complex class boundaries can be more easily separated [35]. This flexibility enables SVM to adapt to diverse text representations and sentiment distributions, making it suitable for different research contexts.

Another advantage of SVM is its strong performance even with relatively small training datasets compared to other machine learning algorithms [36]. This characteristic is useful in sentiment analysis, where labeled data are often limited or costly to obtain. Due to its theoretical foundation and consistent results, SVM remains a reliable benchmark method in sentiment analysis and broader NLP research. Furthermore, linear, polynomial, and radial basis kernel functions allow SVM to handle nonlinear data patterns effectively. With tweet data, which are typically short and information-dense, SVM has demonstrated high performance in previous studies [37]. Therefore, SVM is used as a comparison model when evaluating the performance of other algorithms in sentiment classification against DT.

2.7. Naïve Bayes (NB)

NB is commonly used in text mining due to its conceptual simplicity and efficiency in handling large-scale textual data. The algorithm predicts expected outcomes based on patterns observed in previous data through Bayes' Theorem, a probabilistic approach that calculates the likelihood of an event based on prior knowledge of related conditions. In text mining, NB is applied to classify documents into specific categories, such as positive or negative sentiment, by analyzing the distribution of words within each class [38].

In practice, the implementation of NB for text classification involves several structured stages. The process begins with preprocessing, where raw text is prepared through tokenization, stopword removal, and stemming to retain only relevant linguistic features. This

is followed by feature weighting, in which word frequencies are calculated to represent the characteristics of each document. Using labeled training data, the model then estimates the probability of each class based on these features, forming the basis for classifying new, unseen text [39].

Despite its efficiency and structured workflow, NB has conceptual limitations. The assumption that each word contributes independently to classification does not fully reflect the nature of human language, where meaning often arises from word combinations and contextual relationships [40]. Consequently, NB may struggle to interpret complex linguistic patterns such as negation, irony, or implicit sentiment, reducing its effectiveness in nuanced text analysis tasks [41–43].

$$P(A|B) = \frac{(P(B|A) \times P(A))}{P(B)} \quad (1)$$

In text mining, the is commonly written as

$$P(Ci|D) = \frac{(P(D|Ci) \times P(Ci))}{P(D)} \quad (2)$$

Mathematically, NB is grounded in Bayes' Theorem, which calculates the probability of an event based on prior probabilities and observed evidence. In programming applications, the probability of event A given event B is determined by the probability of B occurring when A happens, multiplied by the probability of A, and divided by the probability of B. In text classification, this formulation is adapted to estimate the probability of a class given a document, with the class of highest probability selected as the prediction.

3. Methodology

3.1. Algorithms

We employed a quantitative research approach based on text mining and supervised machine learning techniques to analyze public sentiment toward the use of AI in digital art. Social media data served as the primary source of information, as platforms such as X (Twitter) provide rich and spontaneous expressions of public opinion. The methodology consisted of data collection, text preprocessing, sentiment labeling, feature extraction, sentiment classification, and performance evaluation. A comparative analysis was conducted using NB, SVM, and DT algorithms to determine their effectiveness in classifying sentiment polarity.

The research was designed in a supervised learning framework, in which labeled textual data were used to train and evaluate classification models. A comparative analysis was conducted by implementing three widely used machine learning algorithms in sentiment analysis: NB, SVM, and DT. The workflow began with data crawling from social media, followed by preprocessing to reduce noise and improve data quality. Sentiment labels were then assigned to the dataset, textual features were extracted using statistical weighting techniques, and classification models were developed. The final stage involved evaluating and comparing the performance of the models using standard metrics.

3.2. Data Collection and Process

Data were collected from the X platform through a crawling process implemented in Google Colab. Twitter was chosen due to its widespread use and its role as a platform for public discourse on emerging technologies, including AI and digital art. Tweets containing the keyword 'AI generative art' were retrieved to ensure relevance to the research topic. The data collection period spanned from January 2020 to November 2025, enabling the

analysis of public sentiment across different stages of AI development and adoption. A total of approximately 1737 tweets were obtained as the initial dataset.

Data were preprocessed to prepare the raw textual data for analysis and modeling. Since the collected tweets were written in English, preprocessing focused on standard text normalization procedures. These included case folding to convert all text to lowercase, text cleaning to remove URLs, user mentions, hashtags, emojis, and special characters, and tokenization to split the text into individual words or terms. These steps reduced noise and ensured that the textual data were represented in a consistent and structured format suitable for feature extraction and machine learning classification.

Sentiment labeling was performed using a hybrid approach that combined manual annotation and automated techniques. A subset of 329 tweets was manually labeled into two categories, positive and negative, serving as a gold standard for training and validating the models. To efficiently label the remaining tweets, a lexicon-based sentiment analysis tool was applied. This combination of manual and automated labeling balanced accuracy and scalability, which is particularly important when working with large social media datasets.

3.3. Feature Extraction

Feature extraction was conducted to transform the preprocessed textual data into numerical representations for machine learning algorithms. The Term Frequency–Inverse Document Frequency (TF-IDF) method was employed to generate feature vectors. TF-IDF assigns weights to terms based on their frequency within a tweet and their inverse frequency across the dataset, allowing important words to receive higher weights. This method is widely adopted in sentiment analysis research due to its effectiveness in representing textual information and reducing the impact of common but less informative terms.

4. Results and Discussion

This study was conducted using NB, SVM, and DT algorithms, which were implemented through RapidMiner. The first step involved defining the research topic, namely sentiment analysis toward the use of AI in digital art among users in Singapore and the Philippines. Data were collected from the X platform using relevant keywords such as AI generative art, digital art, and AI, in English for both countries.

4.1. Data Crawling

The data crawling process was carried out through several interconnected stages. The initial stage began with entering the Twitter auth token, which is an authentication code obtained from each user's Twitter account. This token serves as access, allowing the account to be used in the tweet collection process. The next stage involved configuring the data search parameters. The search keyword parameter was used to specify the keywords to be searched, such as AI generative art and digital art, so that only tweets containing these keywords would be collected. Next, the since and until parameters were used to limit the time range of the tweets collected, ensuring that the data matched the research period, from January 2020 to November 2025. Additionally, the parameter was applied to filter tweets based on the English language, making the data relevant for Singapore and the Philippines.

To control the number of tweets collected, the limit parameter was set as the maximum number of tweets to be retrieved in a single crawling process. After the data were successfully collected, the next step was initial data processing using the pandas library, using the command import pandas as pd. This library facilitates data cleaning, grouping, and storage in the comma-separated value format, preparing the data for further analysis.

https://doi.org/10.3390/engproc2026141013

The main stages of the data collection process are summarized as follows.

- Account authentication: Enter the Twitter auth token obtained from the user's account to authenticate and enable data crawling.
- Keyword selection: Set the search_keyword to specify the topic of tweets to be collected, such as AI generative art and digital art.
- Time range selection: Configure the since and until parameters to limit the time range of collected tweets according to the research period (January 2020–November 2025).
- Language selection: Use the lang parameter to filter tweets in English to ensure relevance for Singapore and the Philippines.
- Data limit: Set the limit as the maximum number of tweets to be collected in a single crawling session.
- Initial data processing: Use the pandas library to manage the collected data, including cleaning and grouping.
- Data storage: Save the collected data in CSV format as a basis for further analysis.

To implement the data collection process, a series of commands were executed in a notebook environment, as illustrated in Figures 1 and 2. Figure 1 shows the environment setup stage, including the installation of required Python ver 3.9 libraries such as pandas, Node.js dependencies, and Playwright for browser automation. This step ensures that the system is properly configured to perform tweet crawling.

Figure 1. Python code for installing required libraries and setting up the crawling environment, including pandas, Node.js dependencies, and Playwright for browser automation.

Figure 2. Crawling data Python code.

The code in Figure 1 installs the tweet-harvest library and its dependencies, which enable automated data collection from the X platform. The Playwright library is configured to handle browser-based interactions required for authentication and data retrieval.

Figure 2 presents the data crawling and initial processing workflow. In this stage, the Twitter authentication token is defined, followed by specifying the search parameters such as keywords, time range, language, and data limit. The crawling process is executed using the tweet-harvest tool, and the collected data are stored in CSV format. Subsequently, the pandas library is used to read, display, and compute the number of collected tweets, enabling preliminary data inspection before further analysis.

As shown in Figure 2, the Twitter authentication token is first defined to authorize data access. Search parameters such as keywords ('AI generative art', 'digital art'), time range (January 2020 to November 2025), language (English), and data limit are then specified. The tweet-harvest tool executes the crawl, and collected data are saved as a CSV file. The pandas library is subsequently used to read and display the dataset, confirming the total number of tweets collected.

Table 1 presents sample tweets from the raw dataset collected from the X platform, illustrating the variety of user expressions, including opinions on AI-generated art, personal experiences, and discussions about digital creativity.

Table 1. Data set.

@CBPostNSullivan @AndyMasley What if the artist uses more energy making the art at their digital workstation because the process takes so much longer than AI? Isn't it even more lazy to pay someone else to make it than to use AI? Why not tell these people
Let's talk about my three top projects on @monad. @cultverse_ai is building a vibrant digital universe where culture tech and creativity come alive. @thedaks_png is building a vibrant community on Monad it's a digital art space where creativity meets connection
I ve noticed that myself These AI bros claiming to be established artists long before the advent of commercial generative AI never do provide proof do they Me I ve shared art I did back in the 1980s I ve shared images of my illustrations in print going back to 2007
Hiya! So my trainer drew this for Agnes Digital cause she didn t like that she got scammed and was given AI art (cause that s bad eww). She loved this Strawberry Lolita dress art the most and it s a shame it wasn t legit so she drew that! Anyways Isn t Digitan cute https://t.co/OeJDbkwg4u (accessed on 4 January 2026)
#DAY6 handmade digital art @gensynai is turning idle GPUs and CPUs into a global supercomputer to train the world's biggest AI models. The era of permissionless AI infrastructure is here. @fenbielding https://t.co/I5RWakfyYN (accessed on 4 January 2026)
@CBPostNSullivan @AndyMasley What if the artist uses more energy making the art at their digital workstation because the process takes so much longer than AI? Isn't it even more lazy to pay someone else to make it than to use AI? Why not tell these people

4.2. Data Preparation

At this stage, the crawled data is processed so that it is ready for analysis using the NB classifier, SVM, and DT algorithms (Figure 3). Preprocessing is carried out through several interrelated steps, including data cleaning, tokenization, and feature transformation.

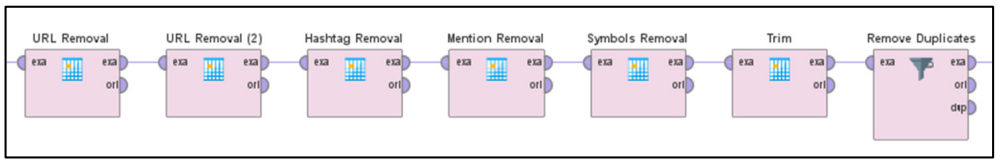


Figure 3. Data cleaning process in Rapid Miner.

Data cleaning is performed to remove noise and irrelevant information from the raw tweets. This step includes eliminating URLs, mentions (@user), hashtags (#topic), punctuation, numbers, and special characters. In addition, all text is converted to lowercase

to ensure consistency, and duplicate or empty tweets are removed. This process helps improve the quality of the dataset and ensures that only meaningful textual information is retained for analysis.

After cleaning, tokenization is applied to split the text into individual words or tokens. These tokens are then transformed into numerical features using appropriate techniques, making them suitable for classification using NB, SVM, and DT algorithms.

Figure 3 illustrates the preprocessing pipeline in RapidMiner. The workflow begins with reading the raw CSV data, followed by a series of cleaning operators including URL removal, mention removal, hashtag removal, symbol removal, trimming, duplicate removal, tokenization, case transformation, stopword filtering, and token length filtering. Each operator prepares the text for subsequent feature extraction and classification.

4.2.1. URL Removal

This step aims to remove uniform resource locators (URLs) from the text, as URLs have no meaning in sentiment analysis (Table 2).

Table 2. URL removal results.

Text	Text After URL Removal
Hiya! So my trainer drew this for Agnes Digital cause she didn t like that she got scammed and was given AI art (cause thats bad eww). She loved this Strawberry Lolita dress art the most and its a shame it wasn t legit so she drew that! Anyways Isnt Digitan cute https://t.co/OeJDbkwg4u (accessed on 4 January 2026)	Hiya! So my trainer drew this for Agnes Digital cause she didn't like that she got scammed and was given AI art (cause thats bad eww). She loved this Strawberry Lolita dress art the most and its a shame it wasn t legit so she drew that! Anyways Isnt Digitan cute

4.2.2. Mention Removal

This step is used to remove mentions (@username) so that the analysis focuses only on the content of the tweet (Table 3).

Table 3. Mention removal results.

Text	Text After Mention Removal
@CBPostNSullivan @AndyMasley What if the artist uses more energy making the art at their digital workstation because the process takes so much longer than AI? Isn't it even more lazy to pay someone else to make it than to use AI? Why not tell these people	What if the artist uses more energy making the art at their digital workstation because the process takes so much longer than AI? Isn't it even more lazy to pay someone else to make it than to use AI? Why not tell these people

4.2.3. Hashtag Removal

This step removes hashtags from the text so that they do not affect keyword analysis (Table 4).

Table 4. Hashtag removal results.

Text	Text After Hashtag Removal
#DAY6 handmade digital art @gensynai is turning idle GPUs and CPUs into a global supercomputer to train the world's biggest AI models. The era of permissionless AI infrastructure is here. @fenbielding https://t.co/I5RWakfyYN (accessed on 4 January 2026)	handmade digital art @gensynai is turning idle GPUs and CPUs into a global supercomputer to train the world's biggest AI models. The era of permissionless AI infrastructure is here. @fenbielding https://t.co/I5RWakfyYN

4.2.4. Symbols Removal

This step removes symbols so that the text is cleaner and easier to process (Table 5).

Table 5. Symbols removal results.

Text	Text After Symbol Removal
@CBPostNSullivan @AndyMasley What if the artist uses more energy making the art at their digital workstation because the process takes so much longer than AI? Isn't it even more lazy to pay someone else to make it than to use AI? Why not tell these people	@CBPostNSullivan @AndyMasley What if the artist uses more energy making the art at their digital workstation because the process takes so much longer than AI Isn't it even more lazy to pay someone else to make it than to use AI Why not tell these people

4.2.5. Trimming

Trimming is conducted to remove unnecessary spaces at the beginning and end of text (Table 6).

Table 6. Trimming results.

Text	Text After Trimming
I ve noticed that myself These AI bros claiming to be established artists long before the advent of commercial generative AI never do provide proof do they Me I ve shared art I did back in the 1980s I ve shared images of my illustrations in print going back to 2007	I ve noticed that myself These AI bros claiming to be established artists long before the advent of commercial generative AI never do provide proof do they Me I ve shared art I did back in the 1980s I ve shared images of my illustrations in print going back to 2007

4.2.6. Remove Duplicates

Duplicate removal ensures that each tweet is unique so that the model results are not biased due to data duplication (Table 7).

Table 7. Remove duplicate results.

Text	Text After Duplicate Removal
– yumayanagisawa Hi I m a Happy to connect creative developer with experience in TouchDesigner Unreal Engine and generative AI I ve worked on interactive installations and media art projects and I d love to explore a project based collaboration and discuss details – yumayanagisawa Hi I m a creative developer with experience in TouchDesigner Unreal Engine and generative AI I ve worked on interactive installations and media art projects and I d love to explore a project based collaboration Happy to connect and discuss details – Kosti_box Astrel_Ninja jonah hill is the guy in this reaction meme and i dont consider a picture from the movie being used as a reaction meme as art as pretentious as it sounds So using an ai baby in similar vein conveying a reaction isnt intruding into art with generative ai t co mcqo9T0v2T – Kosti_box Astrel_Ninja im against generative ai in artistic spaces Not necessarily against all of AI You d have to convince me that the baby constitutes as ai art Not even trying to be obtuse here i just dont see it that way The picture of jonah hill wouldnt be art no	– yumayanagisawa Hi I m a creative developer with experience in TouchDesigner Unreal Engine and generative AI I ve worked on interactive installations and media art projects and I d love to explore a project based collaboration Happy to connect and discuss details – Kosti_box Astrel_Ninja jonah hill is the guy in this reaction meme and i dont consider a picture from the movie being used as a reaction meme as art as pretentious as it sounds So using an ai baby in similar vein conveying a reaction isnt intruding into art with generative ai t co mcqo9T0v2T

4.2.7. Tokenization

Tokenization divides sentences or documents into word units so that the text can be processed and analyzed (Table 8).

Table 8. Tokenization results.

Text	Text After Tokenization
I hate how generative AI is ruining art and writing	I hate how generative AI is ruining art and writing

4.2.8. Transformation

The transform case function is used to convert all text to lowercase to avoid variations in uppercase and lowercase letters that affect the analysis (Table 9).

Table 9. Transform code removal results.

Text	Text After Transformation
I'm going to keep saying it GENERATIVE AI is not art Those who are shoving money into it want to convince you that they can make art But the purpose of art is the antithesis of the goal of Gen AI And that s what makes art so beautiful It can never be imitated	i m going to keep saying it generative ai is not art those who are shoving money into it want to convince you that they can make art but the purpose of art is the antithesis of the goal of gen ai and that s what makes art so beautiful it can never be imitated

4.2.9. Filter Stopwords

Stopwords are common words that have no significant meaning, such as "i" and "and that", so they need to be removed so that the analysis focuses on meaningful words (Table 10).

Table 10. Filter stopword removal results.

Text	Text After Stopword Removal
i miss the times where generative ai wasn t that advanced and normalized but at the same time it has brought more value to actual human made art no matter how messy or crappy it looks	miss the times where generative ai wasn t advanced normalized but at the same time it has brought more value to actual human made art no matter how messy or crappy it looks

4.2.10. Filter Token by Length

This filter filters out irrelevant tokens, such as numbers, symbols, or URLs, so that only meaningful tokens are used in the analysis (Table 11).

Table 11. Filter Token by Length.

Text	Text After Filtering Token Length
Remember it s always ethically correct to bully generative AI artists and empower them to pick up a pencil or some digital art software t co n	Remember always ethically correct bully generative artists empower them pick up pencil some digital software

4.3. Modelling

After completing the data cleaning process through the various removal steps described earlier, the next stage involved modeling the dataset to classify tweets into positive and negative sentiments.

4.3.1. Manual Sentiment Labeling

Before conducting automated sentiment analysis using RapidMiner, manual labelling was applied to the initial 329 tweets (Figure 4). Positive labels were assigned to tweets that conveyed helpfulness or ease, while negative labels were assigned to tweets that contained

criticism or negative expressions. The purpose of this manual stage is to provide initial training for the system. With manually labelled data, RapidMiner can learn the pattern of sentiment labelling so that when automated analysis is applied to a larger dataset, the software can classify sentiment more accurately. Once manual labelling was completed, the data were stored in RapidMiner to be used for training the model.

full_text	Sentiment
The Word After Us was my first generative art AI poetry collab with sashastiles The installation version is summary of our work hand made paper as analog genera	Negative
PsyopAnime in animation EVERY FRAME has a creative process and intention behind generative ai is trash and will never get anything remotely similar to human art	Negative
just 2 b clear generative AI is corrosive to the human brain amp; society at large a disgusting extractive tool of modern techno capitalism amp; there is no place for	Positive
i predict 2026 will be the year where the vibe on AI art shifts not because fully AI generated stuff is going to become any less soulless but because human artists are goir	Negative
in my coldest take ai art really just rebranded slop along with other generative content shouldnt exist or at the very least be open to the public and only kept for big c	Positive
7 Finance Explainers stocks cryptos budgeting 8 AI Art amp; Generative Creations 9 Tech Reviews amp; Unboxings 10 Book Summaries amp; Reviews	Positive
cassidyiscringe slimeington chumblings Was ethically crafted using paid VAs and current devs at Embark to train the voice changer No generative AI is used for any	Positive
While other studios face backlash for AI generated art and voices Ubisoft s Teammates experiment shows what happens when you use generative AI to enhance gam	Positive
pavanHQ yuruyurau That looks like a generative art animation where code snippets about Groq s AI tech like LPU and inference morph into colorful glowing explosi	Negative

Figure 4. Manual sentiment result.

As depicted in Figure 4, each tweet in the manually labeled subset was reviewed and assigned either a 'positive' or 'negative' label. Positive labels were given to tweets expressing support, appreciation, or perceived benefits of AI in digital art, while negative labels were assigned to tweets containing criticism, concern, or rejection of AI-generated art. This labeled dataset served as the ground truth for training the NB, SVM, and DT classifiers.

4.3.2. Data Training in RapidMiner

The data storage stage prepares the 329 manually labelled tweets as the basis for initial model training. The process includes the following steps:

- Using the nominal-to-text operator to convert attributes from nominal to categorical;
- Applying filters to select sentiment data that is not missing (is not missing), ensuring only manually labelled data is used;
- Using process documents to transform raw text into numerical features suitable for the NB, SVM, and DT algorithms;
- The NB, SVM, and DT operators are used to train the classification model based on features generated from process documents;
- After training, store model and store data are used to save the trained model and the manually labelled dataset, making it ready for automated sentiment analysis on the remaining unlabelled data.

The workflow in Figure 5 begins with reading the manually labeled dataset. The 'Filter Examples' operator selects only tweets with non-missing sentiment labels. 'Nominal to Text' converts categorical attributes into text format. The 'Process Documents' operator performs tokenization, case conversion, stopword removal, and TF-IDF transformation. The Naïve Bayes operator then trains the classification model, which is saved using 'Store Model' for later application to unlabeled data.

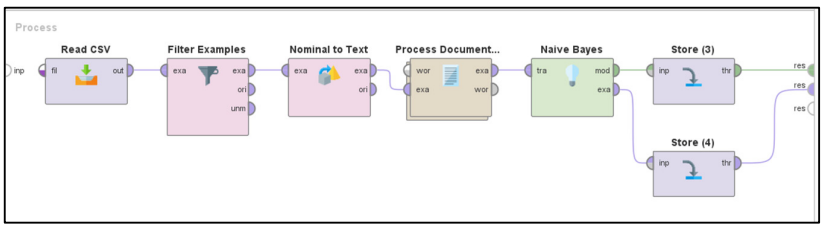


Figure 5. NB for training in RapidMiner.

Figure 5 shows the RapidMiner training workflow for the Naïve Bayes classifier. The workflow begins by loading the manually labeled dataset of 329 tweets. The "Process

Documents" operator performs text preprocessing, including tokenization, case conversion, stopword removal, and token length filtering. Preprocessed text is then converted into numerical feature vectors using TF-IDF. The "Naïve Bayes" operator applies Bayes. Theorem to calculate the probability of positive or negative sentiment based on word distributions. Finally, the trained model is saved using "Store Model" for later application to unlabeled data.

Figure 6 illustrates the SVM classification workflow, starting from reading the dataset and filtering relevant data, followed by data transformation and text preprocessing. The processed data are then classified using the SVM algorithm, and the results are stored for further analysis. For Figures 6 and 7, the SVM/DT training workflow follows the same structure as Figure 5, with the respective algorithm operator substituted for Naïve Bayes.

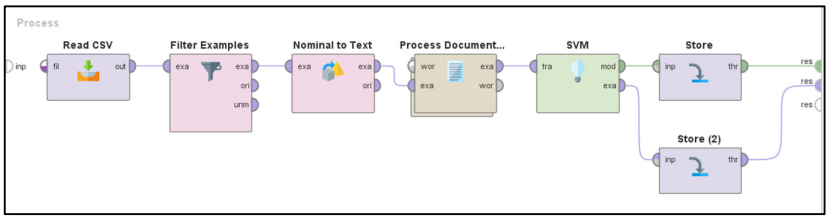


Figure 6. SVM for training in RapidMiner.

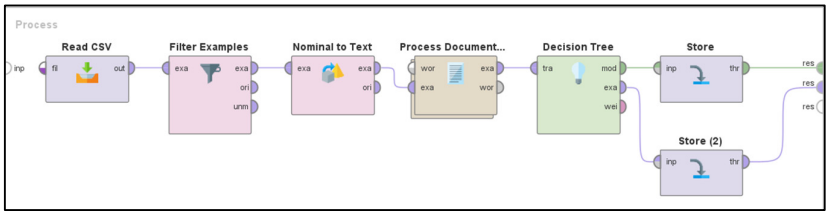


Figure 7. DT for training in RapidMiner.

Figure 7 illustrates the DT classification workflow, starting from reading the dataset and filtering relevant data, followed by data transformation and text preprocessing. The processed data are then classified using the DT algorithm, and the results are stored for further analysis.

4.3.3. Data Test in RapidMiner

Figures 8 and 9 show the sentiment classification data results of the testing data. The model classified 530 tweets as positive sentiment and 1207 tweets as negative sentiment. This result indicates that negative sentiment toward AI in digital art is more dominant among users, although positive opinions remain present.

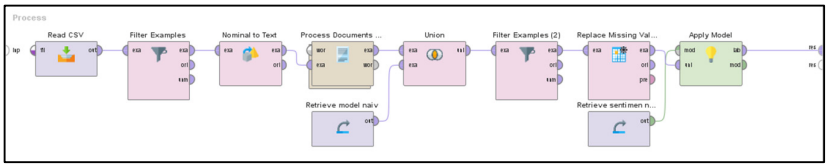
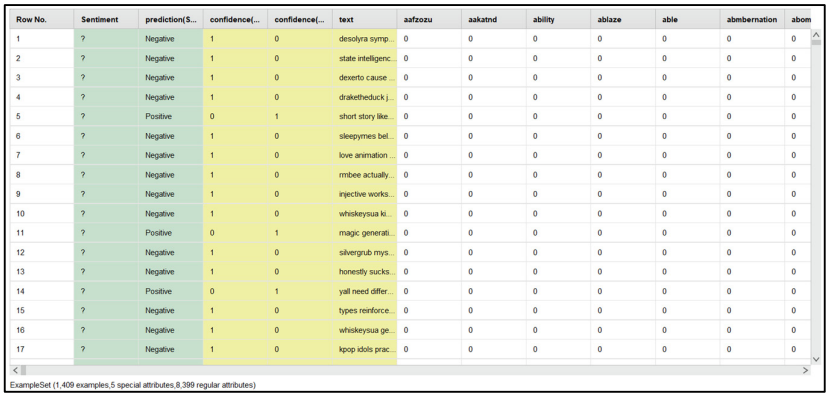


Figure 8. NB for test in RapidMiner.

As shown in Figure 8, the testing workflow reads the full dataset (both labeled and unlabeled), filters to select only tweets without manual labels, applies the same text preprocessing and TF-IDF transformation used during training, and then applies the saved Naïve Bayes model to predict sentiment. Results are stored for analysis.



Row No.	Sentiment	prediction(S...	confidence(...	confidence(...	text	aafzozu	aakatnd	ability	ablaze	able	abmbernation	abom
1	?	Negative	1	0	desolyra symp...	0	0	0	0	0	0	0
2	?	Negative	1	0	state intelligenc...	0	0	0	0	0	0	0
3	?	Negative	1	0	dexerto cause ...	0	0	0	0	0	0	0
4	?	Negative	1	0	drakethedeuck j...	0	0	0	0	0	0	0
5	?	Positive	0	1	short story like...	0	0	0	0	0	0	0
6	?	Negative	1	0	sleepyrnes bel...	0	0	0	0	0	0	0
7	?	Negative	1	0	love animation ...	0	0	0	0	0	0	0
8	?	Negative	1	0	rmbee actually...	0	0	0	0	0	0	0
9	?	Negative	1	0	injective works...	0	0	0	0	0	0	0
10	?	Negative	1	0	whiskeysua ki...	0	0	0	0	0	0	0
11	?	Positive	0	1	magic generati...	0	0	0	0	0	0	0
12	?	Negative	1	0	silvergrub mys...	0	0	0	0	0	0	0
13	?	Negative	1	0	honestly sucks...	0	0	0	0	0	0	0
14	?	Positive	0	1	yall need differ...	0	0	0	0	0	0	0
15	?	Negative	1	0	types reinforce...	0	0	0	0	0	0	0
16	?	Negative	1	0	whiskeysua ge...	0	0	0	0	0	0	0
17	?	Negative	1	0	kpop idols prac...	0	0	0	0	0	0	0

ExampleSet (1,409 examples,5 special attributes,8,399 regular attributes)

Figure 9. Prediction result NB in RapidMiner.

Figures 10 and 11 present the sentiment classification data and results obtained using the SVM model. The figures show that the majority of tweets in the testing dataset were classified as negative sentiment, with 1621 tweets labeled as negative and only 116 tweets classified as positive. This result indicates that the SVM model exhibited a strong bias toward the negative class, resulting in a highly imbalanced sentiment distribution.

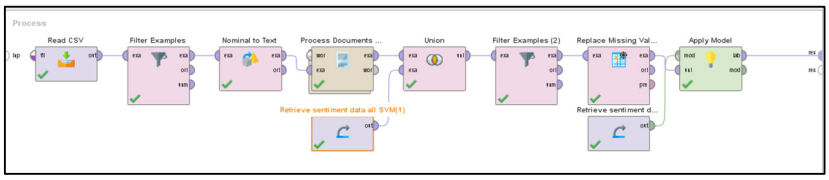
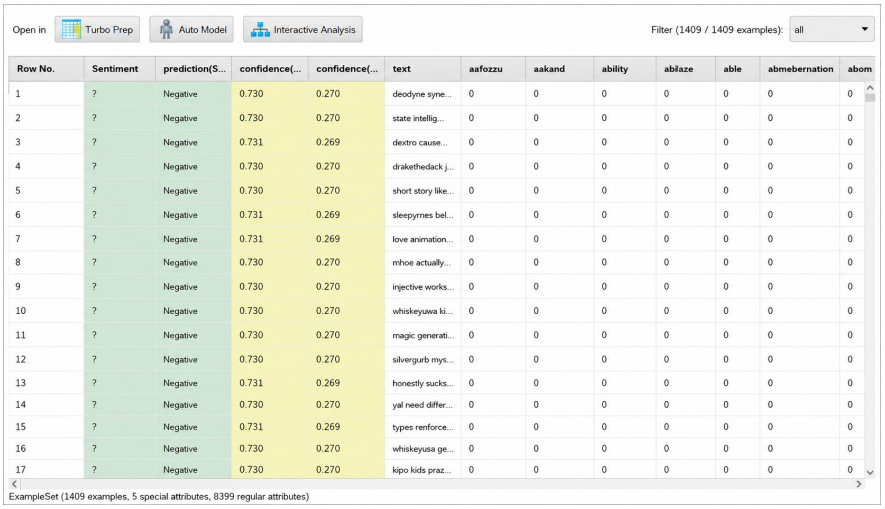


Figure 10. SVM for test in RapidMiner.



Open in Turbo Prep Auto Model Interactive Analysis Filter (1409 / 1409 examples): all

Row No.	Sentiment	prediction(S...	confidence(...	confidence(...	text	aafozzu	aakand	ability	ablaze	able	abmebernation	abom
1	?	Negative	0.730	0.270	deodyne syne...	0	0	0	0	0	0	0
2	?	Negative	0.730	0.270	state intellig...	0	0	0	0	0	0	0
3	?	Negative	0.731	0.269	dextro cause...	0	0	0	0	0	0	0
4	?	Negative	0.730	0.270	drakethedack j...	0	0	0	0	0	0	0
5	?	Negative	0.730	0.270	short story like...	0	0	0	0	0	0	0
6	?	Negative	0.731	0.269	sleepyrnes bel...	0	0	0	0	0	0	0
7	?	Negative	0.731	0.269	love animation...	0	0	0	0	0	0	0
8	?	Negative	0.730	0.270	rmhoe actually...	0	0	0	0	0	0	0
9	?	Negative	0.730	0.270	injective works...	0	0	0	0	0	0	0
10	?	Negative	0.730	0.270	whiskeyuwa ki...	0	0	0	0	0	0	0
11	?	Negative	0.730	0.270	magic generati...	0	0	0	0	0	0	0
12	?	Negative	0.730	0.270	silvergurb mys...	0	0	0	0	0	0	0
13	?	Negative	0.731	0.269	honestly sucks...	0	0	0	0	0	0	0
14	?	Negative	0.730	0.270	yal need differ...	0	0	0	0	0	0	0
15	?	Negative	0.731	0.269	types renforce...	0	0	0	0	0	0	0
16	?	Negative	0.730	0.270	whiskeyusa ge...	0	0	0	0	0	0	0
17	?	Negative	0.730	0.270	kipo kids praz...	0	0	0	0	0	0	0

ExampleSet (1409 examples, 5 special attributes, 8399 regular attributes)

Figure 11. Prediction result data test SVM in RapidMiner.

Figure 10 shows the RapidMiner workflow for applying the trained SVM model to unlabeled test data.

Figure 11 displays the SVM prediction output. Compared to Naïve Bayes, the SVM model produced a much stronger bias toward negative sentiment, classifying only 6.7% of tweets as positive. This suggests that the default SVM parameters may be suboptimal for this imbalanced short-text dataset.

Figure 12 shows the sentiment classification results of the testing data using the DT model. The model classified 195 tweets as positive sentiment and 1542 tweets as negative sentiment, indicating that negative sentiment dominates the dataset.

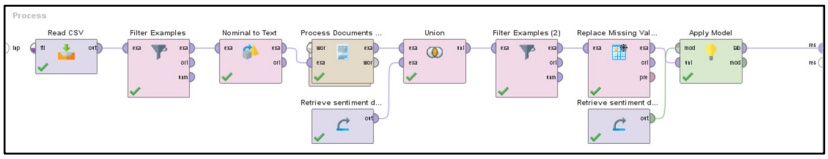


Figure 12. Prediction result data test DT in RapidMiner.

The Decision Tree results in Figure 11 show an intermediate distribution between Naïve Bayes and SVM, with 11.2% positive classification. While less biased than SVM, the Decision Tree still favored negative sentiment, reflecting the underlying class imbalance in the dataset and the model's sensitivity to feature distribution.

Once all the algorithms were trained, the next step was to automatically classify sentiment for tweets without labels. This process involves the following.

1. Read CSV: Loading the cleaned dataset, including manually labelled tweets and the remaining unlabelled tweets.
2. Filter Examples: Selecting only the unlabelled tweets (is missing) so that already labelled data is not processed again.
3. Nominal to Text: Converting categorical attributes into text for further processing.
4. Process Documents: Transforming text into numerical representations using TF-IDF, which evaluates the importance of a word within a tweet relative to the entire dataset.
    - Tokenize: Splitting sentences into individual words.
    - Transform Case: Converting all text to lowercase for consistency.
    - Filter Stopwords: Removing common words with little semantic meaning.
    - Filter Tokens by Length: Removing tokens that are too short or irrelevant.
5. Union: Combining the training and testing datasets so that the model processes all data consistently.
6. Filter Examples (again): Processing the unlabelled data after merging datasets.
7. Replace Missing Value: Filling missing values with zero to prevent model errors during execution.
8. Apply Model: Applying the trained NB, SVM, and DT models to the testing data to predict positive and negative sentiment.

Through these steps, the entire dataset can be analyzed automatically based on sentiment, enabling the NB, Support Vector Machine (SVM), and DT models to efficiently and accurately classify tweets as positive or negative.

4.4. Results

Figure 9 presents the sentiment classification results of the first model. The model identified 530 tweets (≈30.5%) as positive and 1207 tweets (≈69.5%) as negative. These findings suggest that negative sentiment toward AI in digital art is more prevalent, although a notable portion of positive sentiment remains, reflecting perceptions of AI-generated art as beneficial for creativity and efficiency. Figure 11 shows the results obtained using the SVM model. The majority of tweets were classified as negative, with 116 (≈6.7%) labeled positive and 1621 (≈93.3%) labeled negative. This outcome indicates that the SVM model displayed a strong bias toward the negative class, highlighting a limitation of SVM in short-text sentiment analysis when parameter tuning or data balancing is insufficient. Figure 12 illustrates the classification results of the DT model. It identified 195 tweets (≈11.2%) as positive and 1542 (≈88.8%) as negative. While DT captured more positive sentiment than SVM, negative sentiment still dominated, suggesting that the model was also affected by data imbalance and feature distribution, which limited its ability to separate classes effectively.

Collectively, the comparison across Figures 9, 11 and 12 show that different algorithms produce varying sentiment distributions. Although all models reveal a predominance of negative sentiment, the degree of bias differs, underscoring the importance of algorithm selection and parameter optimization in sentiment analysis.

The comparative results demonstrate that NB produced a more balanced sentiment distribution compared with SVM and DT (Tables 12 and 13). Meanwhile, SVM and DT tended to favor the negative class, indicating potential limitations related to parameter configuration and data imbalance. These findings suggest that NB is more suitable for sentiment analysis of short social media texts in this study, while SVM and DT require further optimization to achieve comparable performance.

Table 12. Algorithm comparison of sentiment results.

Algorithm	Positive	Positive (%)	Negative	Negative (%)
NB	530	30.5%	1207	69.5%
SVM	116	6.7%	1621	93.3%
DT	195	11.2%	1542	88.8%

Table 13. Advantages and limitations of classification algorithms.

Algorithm	Advantages	Limitations
NB	Simple and computationally efficient, performs well on short-text data, relatively robust to noise	Assumes feature independence, which may not always reflect real language patterns
SVM	Effective in high-dimensional spaces, theoretically strong generalization capability	Highly sensitive to parameter settings and class imbalance, prone to classification bias in short-text sentiment analysis
DT	Easy to interpret, capable of modeling non-linear relationships	Prone to overfitting, performance strongly depends on feature distribution and tree depth

As shown in Table 12, the terms “positive” and “negative” refer to the sentiment classification results of user tweets related to digital art on X. Positive sentiment indicates that a tweet expresses favorable opinions or support toward digital art, whereas negative sentiment reflects criticism or unfavorable opinions. The “Positive” and “Negative” columns represent the total number of tweets classified into each sentiment category by each algorithm. Meanwhile, “Positive (%)” and “Negative (%)” indicate the proportion of tweets in each category relative to the total dataset, expressed as percentages, allowing for easier comparison of classification outcomes across algorithms.

Table 13 summarizes the key advantages and limitations of the Naïve Bayes, SVM, and Decision Tree algorithms based on their performance in this sentiment analysis study.

5. Conclusions

Public sentiment toward the use of AI in digital art on the X platform is predominantly negative. Among the three classification algorithms applied, NB produced the most balanced results, classifying 530 tweets as positive and 1207 as negative. In contrast, SVM and DT showed a strong bias toward negative sentiment, identifying only 116 and 195 positive tweets, respectively. These outcomes suggest that NB is more suitable for sentiment analysis of short and sparse social media texts. The performance limitations observed in SVM and DT highlight challenges related to class imbalance and feature representation. Future research may therefore focus on optimizing SVM parameters, applying data balancing techniques, and exploring advanced feature extraction or deep

learning models to improve sentiment classification performance in the context of AI-generated digital art.

Author Contributions: Conceptualization, R.M.S., V.C.W., J.A.G., K.G. and I.M.S.; methodology, V.C.W., J.A.G., V.T., F.P.S.S., and K.G.; software, V.C.W. and I.M.S.; validation, R.M.S.; formal analysis, V.C.W., J.A.G. and K.G.; investigation, V.C.W. and I.M.S.; resources, V.C.W. and I.M.S.; data curation, V.C.W. and D.A.A.L.; writing original draft preparation, V.C.W.; writing review and editing, J.A.G., K.G. and I.M.S.; visualization, V.C.W., J.A.G. and I.M.S.; supervision, R.M.S.; project administration, R.M.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data is unavailable due to privacy.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. AlBadani, B.; Shi, R.; Dong, J. A Novel Machine Learning Approach for Sentiment Analysis on Twitter Incorporating the Universal Language Model Fine-Tuning and SVM. *Appl. Syst. Innov.* **2022**, *5*, 13. [CrossRef]
2. Amro, A.; Al-Akhras, M.; El Hindi, K.; Habib, M.; Abu Shawar, B. Instance Reduction for Avoiding Overfitting in Decision Trees. *J. Intell. Syst.* **2021**, *30*, 438–459. [CrossRef]
3. Asyaky, M.S.; Al-Husaini, M.; Lukmana, H.H. Sentiment Analysis on Short Social Media Texts Using DistilBERT. *J. Comput. Netw. Archit. High Perform. Comput.* **2025**, *7*, 524–533. [CrossRef]
4. Bolaj, K. Text Categorization System for English Text Documents using NB Classifier. *Int. J. Comput. Appl.* **2020**, *177*, 7–10.
5. Bonacchi, C.; Krzyzanska, M.; Acerbi, A. Positive sentiment and expertise predict the diffusion of archaeological content on social media. *Sci. Rep.* **2025**, *15*, 11234. [CrossRef] [PubMed]
6. Chakraborty, K.; Bhattacharyya, S.; Bag, R. A Survey of Sentiment Analysis from Social Media Data. *IEEE Trans. Comput. Soc. Syst.* **2020**, *7*, 450–464. [CrossRef]
7. Jijo, B.T.; Abdulazeez, A.M. Classification Based on Decision Tree Algorithm for Machine Learning. *J. Appl. Sci. Technol. Trends* **2021**, *2*, 20–28. [CrossRef]
8. Custode, L.L.; Iacca, G. Evolutionary Learning of Interpretable Decision Trees. *IEEE Access* **2020**, *11*, 2169–3536. [CrossRef]
9. Dervenis, C.; Kanakis, G.; Fitsilis, P. Sentiment analysis of student feedback: A comparative study employing lexicon and machine learning techniques. *Stud. Educ. Eval.* **2024**, *83*, 101406. [CrossRef]
10. Du, K.-L.; Jiang, B.; Lu, J.; Hua, J.; Swamy, M.N.S. Exploring Kernel Machines and Support Vector Machines: Principles, Techniques, and Future Directions. *Mathematics* **2024**, *12*, 3935. [CrossRef]
11. Epstein, Z.; Hertzmann, A. The Investigators of Human Creativity. Art and the science of generative AI. *Science* **2023**, *380*, 1110–1111. [CrossRef]
12. Feng, W.; Li, Y.; Ma, C.; Yu, L. From ChatGPT to Sora: Analyzing Public Opinions and Attitudes on Generative AI in Social Media. *IEEE Access* **2025**, *13*, 14485–14498. [CrossRef]
13. Feuerriegel, S.; Hartmann, J.; Janiesch, C.; Zschech, P. Generative AI. *Bus. Inf. Syst. Eng.* **2023**, *66*, 111–126. [CrossRef]
14. Fu, Y.; Bin, H.; Zhou, T.; Wang, M. Creativity in the Age of AI: Evaluating the Impact of Generative AI on Design Outputs and Designers' Creative Thinking. *arXiv* **2024**, arXiv:2411.00168. [CrossRef]
15. Mukherjee-Gandhi, A.; Muellerklein, O. When Algorithms Meet Artists: Topic Modeling the AI-Art Debate, 2013–2025. *arXiv* **2025**, arXiv:2508.03037. [CrossRef]
16. Han, Y.; Yu, J.; Zhang, N.; Meng, C.; Ma, P.; Zhong, W.; Zou, C. Leverage Classifier: Another Look at Support Vector Machine. *Stat. Sin.* **2023**, *33*, 1605–1625. [CrossRef]
17. Hassani, H.; Beneki, C.; Unger, S.; Mazinani, M.T.; Yeganegi, M.R. Text Mining in Big Data Analytics. *Big Data Cogn. Comput.* **2020**, *4*, 1. [CrossRef]
18. Ma, J.; Fan, L.; Tian, W.; Miao, Z. Research on Data Classification Method of Optimized Support Vector Machine Based on Gray Wolf Algorithm. *Int. J. Grid High Perform. Comput.* **2023**, *15*, 1–14. [CrossRef]
19. Jukanti, M. Understanding Natural Language Processing (NLP) Techniques. *J. Comput. Sci. Technol. Stud.* **2025**, *7*, 1005–1012. [CrossRef]

20. Punitha, K.; Raja Shree, S.; Cruz Antony, J. A Systematic Review of Sentiment Analysis Approaches and Techniques. In Proceedings of the 2025 6th International Conference on Mobile Computing and Sustainable Informatics (ICMCSI), Goathgaun, Nepal, 7–8 January 2025; pp. 603–607.
21. Klusowski, J.M.; Tian, P.M. Large Scale Prediction with DTs. *J. Am. Stat. Assoc.* **2021**, *119*, 525–537. [CrossRef]
22. Kumar, R.; Goswami, B.K.; Mhatre, S.M.; Agrawal, S. Naive Bayes in Focus: A Thorough Examination of its Algorithmic Foundations and Use Cases. *Int. J. Innov. Sci. Res. Technol.* **2024**, *9*, 2078–2081. [CrossRef]
23. Fitrana, L.A.; Linawati, S.; Herlinawati, N.; Sa'adah, R.; Seimahuria, S. Analysis of Twitter User Sentiment Towards the Indosat Brand Using the Naïve Bayes Classifier Method (Analisis Sentimen Pengguna Twitter Terhadap Brand Indosat Menggunakan Metode Naïve Bayes Classifier). *J. Mhs. Tek. Inform.* **2024**, *8*, 4291–4297. (In Indonesian)
24. Lee, K.-P.; Song, S. Developing insights from the collective voice of target users in Twitter. *J. Big Data* **2022**, *9*, 45. [CrossRef]
25. Pradeep, M.; Sasivardhan, T.; Bodana, G.; Shilpa, K.; Savalapurapu, K.; Babu, G.C. Natural Language Processing for Literacy Text Mining: Extracting Knowledge From British National Corpus. In Proceedings of the 2025 6th International Conference on Inventive Research in Computing Applications (ICIRCA), Coimbatore, India, 25–27 June 2025; pp. 1816–1821.
26. Martín, M.S.; Chen, F.-W.; Urbistondo, P.A. Application of the LDA model to identify topics in telemedicine conversations on the X social network. *BMC Health Serv. Res.* **2025**, *25*, 369. [CrossRef] [PubMed]
27. Messer, U. Co-creating art with generative AI: Implications for artworks and artists. *Comput. Hum. Behav. Artif. Hum.* **2024**, *2*, 100056. [CrossRef]
28. Mujahid, M.; Kına, E.; Rustam, F.; Villar, M.G.; Alvarado, E.S.; Diez, I.D.L.T.; Ashraf, I. Data oversampling and imbalanced datasets: An investigation of performance for machine learning and feature engineering. *J. Big Data* **2024**, *11*, 87. [CrossRef]
29. Musyarofah, A.M.; Prasetyo, R.F. The Perception of Fine Arts Students at Semarang State University towards Artificial Intelligence in the Field of Business and Fine Arts Education to the General Public (Persepsi Mahasiswa Seni Rupa Universitas Negeri Semarang terhadap Kecerdasan Buatan di Ranah Bisnis dan Pendidikan Seni Rupa kepada Masyarakat Umum). *J. Majemuk* **2024**, *3*, 387–400. (In Indonesian)
30. Nordström, P.; Lundman, R.; Hautala, J. Evolving Coagency between Artists and AI in the Spatial Cocreative Process of Artmaking. *Ann. Am. Assoc. Geogr.* **2023**, *113*, 2203–2218. [CrossRef]
31. Ohme, J.; Araujo, T.; Boeschoten, L.; Freelon, D.; Ram, N.; Reeves, B.B.; Robinson, T.N. Digital Trace Data Collection for Social Media Effects Research: APIs, Data Donation, and (Screen) Tracking. *Commun. Methods Meas.* **2023**, *18*, 124–141. [CrossRef]
32. Puadi, M.F.; Hashim, M.E.A.; Albakry, N.S. Between Control and Collaboration: Artistic Autonomy in AI-Generated Visual Artworks. *Semarak Int. J. Creat. Art Des.* **2025**, *4*, 1–11. [CrossRef]
33. Wahyuni, P.; Romli, M.A. Comparison of Naïve Bayes Classifier and Decision Tree Algorithms for Sentiment Analysis on the House of Representatives' Right of Inquiry on Twitter. *J. Appl. Inform. Comput.* **2024**, *8*, 523–530. [CrossRef]
34. Quirita, V.A.A.; Cárdenas, J.D.; Aliaga, W.; Palacios, A.; Sierra, R.B. Peruvian Presidential Debates in the Elections of 2021 in Twitter/X: A Sentiment Analysis Approach. *IEEE Access* **2024**, *12*, 138386–138398. [CrossRef]
35. Rezaeenour, J.; Ahmadi, M.; Jelodar, H.; Shahrooei, R. Systematic review of content analysis algorithms based on deep neural networks. *Multimed. Tools Appl.* **2022**, *82*, 17879–17903. [CrossRef]
36. Stracqualursi, L.; Agati, P. Twitter users perceptions of AI-based e-learning technologies. *Sci. Rep.* **2024**, *14*, 5927. [CrossRef]
37. Tang, J.; Liu, X. 'NO TO AI GENERATED IMAGES': Fan art creators contesting AI integration on social media platforms. *Media Int. Aust.* **2025**, *197*, 137–151. [CrossRef]
38. Uzun, E. A Novel Web Scraping Approach Using the Additional Information Obtained From Web Pages. *IEEE Access* **2020**, *8*, 61726–61740. [CrossRef]
39. Van Atteveldt, W.; Van der Velden, M.A.C.G.; Boukes, M. The Validity of Sentiment Analysis: Comparing Manual Annotation, Crowd-Coding, Dictionary Approaches, and Machine Learning Algorithms. *Commun. Methods Meas.* **2021**, *15*, 121–140. [CrossRef]
40. Van Der Veen, A.M.; Bleich, E. The advantages of lexicon-based sentiment analysis in an age of machine learning. *PLoS ONE* **2025**, *20*, e0313092. [CrossRef]
41. Wan, W.; Huang, R. Deep Learning-Driven Public Opinion Analysis on the Weibo Topic about AI Art. *Appl. Sci.* **2024**, *14*, 3674. [CrossRef]
42. Wankhade, M.; Rao, A.C.S.; Kulkarni, C. A survey on sentiment analysis methods, applications, and challenges. *Artif. Intell. Rev.* **2022**, *55*, 5731–5780. [CrossRef]
43. Zakharia, A.; Hasanah, H.; Srirahayu, A. Public Sentiment Analysis on Twitter About the Use of AI in Digital Art Using CNN (Analisis Sentimen Publik di Twitter Tentang Pemanfaatan AI dalam Seni Digital Menggunakan CNN). *J. Inform. Tek. Elektro Terap.* **2025**, *13*, 536–545. (In Indonesian)