



Article

Comparing the Performance of Machine Learning Algorithms in the Automatic Classification of Psychotherapeutic Interactions in Avatar Therapy

Alexandre Hudon ^{1,2} , Kingsada Phraxayavong ³ , Stéphane Potvin ^{1,2} and Alexandre Dumais ^{1,2,3,4,*}

¹ Department of Psychiatry and Addictology, Faculty of Medicine, Université de Montréal, Montréal, QC H3T 1J4, Canada; alexandre.hudon.1@umontreal.ca (A.H.); stephane.potvin@umontreal.ca (S.P.)
² Centre de Recherche de l'Institut Universitaire en Santé Mentale de Montréal, Montréal, QC H1N 3J4, Canada
³ Services et Recherches Psychiatriques AD, Montréal, QC H1C 1H1, Canada; kingsada@me.com
⁴ Institut National de Psychiatrie Légale Philippe-Pinel, Montréal, QC H1C 1H1, Canada
* Correspondence: alexandre.dumais@umontreal.ca

Abstract: (1) Background: Avatar Therapy (AT) is currently being studied to help patients suffering from treatment-resistant schizophrenia. Facilitating annotations of immersive verbatims in AT by using classification algorithms could be an interesting avenue to reduce the time and cost of conducting such analysis and adding objective quantitative data in the classification of the different interactions taking place during the therapy. The aim of this study is to compare the performance of machine learning algorithms in the automatic annotation of immersive session verbatims of AT. (2) Methods: Five machine learning algorithms were implemented over a dataset as per the Scikit-Learn library: Support vector classifier, Linear support vector classifier, Multinomial Naïve Bayes, Decision Tree, and Multi-layer perceptron classifier. The dataset consisted of the 27 different types of interactions taking place in AT for the Avatar and the patient for 35 patients who underwent eight immersive sessions as part of their treatment in AT. (3) Results: The Linear SVC performed best over the dataset as compared with the other algorithms with the highest accuracy score, recall score, and F1-Score. The regular SVC performed best for precision. (4) Conclusions: This study presented an objective method for classifying textual interactions based on immersive session verbatims and gave a first comparison of multiple machine learning algorithms on AT.

Keywords: artificial intelligence; virtual reality therapy; auditory hallucinations; schizophrenia; psychotherapy; machine learning; algorithms



Citation: Hudon, A.; Phraxayavong, K.; Potvin, S.; Dumais, A. Comparing the Performance of Machine Learning Algorithms in the Automatic Classification of Psychotherapeutic Interactions in Avatar Therapy. *Mach. Learn. Knowl. Extr.* **2023**, *5*, 1119–1131. <https://doi.org/10.3390/make5030057>

Academic Editor: Andreas Holzinger

Received: 31 July 2023

Revised: 22 August 2023

Accepted: 22 August 2023

Published: 24 August 2023



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

1. Introduction

A severe mental disorder such as schizophrenia has a high social burden [1]. The economic burden of schizophrenia in the United States alone reached 155.7 billion dollars in 2013 [2]. The mental state of those suffering from schizophrenia may be disturbed. This disturbance can include delusions and hallucinations, also known as positive symptoms. Patients with schizophrenia are more likely to experience auditory hallucinations [3]. A thorough strategy is therefore required for the treatment of positive symptoms. Psychoeducation is used to explain the diagnosis, and psychopharmacological treatments are added to deal with delusions and hallucinations [4,5]. Despite receiving regular medical treatments, over 25% of individuals still have positive symptoms [6,7]. Antipsychotic drugs and psychotherapy techniques such as family interventions, psychoeducation, and cognitive-behavioral therapy (CBT) are frequently used in the standard of care treatment [8,9].

Novel therapies such as Avatar Therapy (AT) emerged to account for this problem and offer an alternate solution for patients suffering from schizophrenia with refractory auditory hallucinations [10]. This therapy is still being studied to validate its efficiency in reducing patients' refractory auditory hallucinations and assessing their wellbeing. Avatar

Therapy implies the use of a virtual reality headset where the therapists interact with the patient in an immersive environment [11]. In the environment, the therapist animates a visual representation (pre-configured by the patient) of the patient's auditory hallucination. AT was initially developed by Leff et al. (2014) in 2008 [12]. In their first pilot trial for this type of therapy, AT consisted of 7 weeks of therapy (one session per week), comprising six immersive 30 min sessions with the Avatar. This trial enrolled 26 patients, 16 received AT, and they benefited from a significant reduction in the frequency and intensity of their auditory hallucinations [13]. Furthermore, it highlighted a significant reduction in depressive symptoms. In 2016, Craig and al. (2018, trial number: ISRCTN, number 65,314 790) conducted the first single-blind, randomized controlled trial with 150 patients from 18 to 65 years who had received a clinical diagnosis of schizophrenia spectrum and had auditory verbal hallucinations despite continued treatment [14]. These patients were randomly assigned to receive AT or supportive therapy. The main outcome was reduction in auditory verbal hallucinations at 12 weeks on the Psychotic Symptoms Rating Scales Auditory Hallucinations (PSYRATS-AH) [14]. At the Institut Universitaire en Santé Mentale de l'Université de Montréal (IUSMM), an ongoing clinical trial piloted by Dr. Dumais and Dr. Potvin is comparing AT to CBT for patients suffering from schizophrenia with auditory hallucinations under continued treatment. The trial includes 136 participants: 68 undergoing AT and 68 undergoing CBT. While this trial is underway, a one-year pilot randomized comparative trial evaluating the short- and long-term efficacy of VRT over CBT at the IUSMM for this population and assessed 37 patients who undertook AT and 37 who undertook CBT [15]. AT achieved larger effect sizes than CBT on auditory hallucinations for these patients as well as showed significant results on persecutory beliefs and quality of life [15].

While clinical trials are showing promising outcomes regarding the impact of Avatar Therapy (AT) in reducing auditory hallucinations among individuals with schizophrenia, a few studies have attempted to qualitatively assess the verbatims of immersive sessions to gain a deeper understanding of the therapeutic process. Commonly employed techniques for this assessment include content analysis of therapeutic sessions, semi-structured interviews, and questionnaires. However, these methods can be time-consuming, require significant human resources, are susceptible to biases depending on the analytical approach taken, and may be hard to generalize [16]. These biases include misclassification of outcomes, selection biases, and confounding biases [17]. Often, they focus on a limited set of items, which makes it challenging to obtain a comprehensive understanding of the underlying therapeutic process. Qualitative approaches such as phenomenology or grounded theory are often utilized to explore the nuances of therapeutic sessions [18].

In 2018, an initial content analysis of AT was conducted, examining the therapeutic sessions of 12 patients who underwent the therapy [19]. They analyzed up to 84 immersive session verbatims until reaching a saturation point. This analysis revealed five thematic areas that emerged from patients' dialogue with the Avatar: emotional response to voices, beliefs about voices and schizophrenia, self-perceptions, coping mechanisms, and aspirations [19]. These themes provided initial insights into potential therapeutic targets in AT. Building upon this, Beaudoin et al. conducted a subsequent study in 2021, qualitatively assessing 125 therapy verbatims (totaling 1419 min) from 18 patients [20]. The aim was to gain a deeper understanding of the dynamics between the patient and the Avatar. Two major key themes were identified for the Avatar: confrontational techniques (comprising eight sub-themes) and positive techniques (comprising six sub-themes). For the patients, five key themes were identified: self-perceptions, emotional responses, aspirations, coping mechanisms, and beliefs about voices and schizophrenia. These five themes encompassed a total of 14 sub-themes [20]. These qualitative studies contribute to the knowledge of the therapeutic process in AT, shedding light on the interactions between patients and Avatars and identifying key thematic areas that could guide future research and therapeutic interventions. While qualitative data can be informative and extensive in nature, it lacks

the quantitative counterpart necessary to determine the specific elements of therapy that may contribute to positive outcomes.

Classification algorithms are often used in the field of medicine to account for this lack of quantitative assessment [21]. As an example, a study designed by Chekroud et al. reviewed the use of classification algorithms to predict treatment outcomes in psychiatry, ranging from medication to psychotherapies to digital interventions and neurobiological treatments, and included the classification of text entities [22]. They conclude that the use of classification algorithms is a new but important approach to improving the effectiveness of mental health care [22]. In mental health, few of these approaches have been attempted, mostly due to the limited amount of data available (e.g., a small number of therapeutic verbatims). In Avatar Therapy, the complexity of having interactions between three individuals and the fact that it is less readily available to the public limits the extent of usable data for constructing a database. As an example, this can yield databases that are smaller than data readily available for internet-based CBT. A classification algorithm applicable to small databases is therefore needed for such cases. A recent review assessed machine learning algorithms used in the context of psychiatry, psychology, and social sciences and identified several potential algorithms that can be used with small datasets [23]. Classification algorithms such as Naïve Bayes, Decision Tree, and support vector machine classifiers were found to be relevant in these contexts. According to the identified algorithms, the most used and best-performing algorithm is the support vector machine [23]. This opens the door to merging previous content analysis with quantifiable data to forecast the prediction of therapeutic outcomes in the context of psychotherapy. Facilitating annotations of immersive verbatims in AT by using classification algorithms could be an interesting avenue to reduce the time and cost of conducting such analysis and adding objective quantitative data in the identification and classification of the different interactions taking place during the therapy.

The aim of this study is to compare the performance of machine learning algorithms in the automatic annotation of immersive session verbatims of AT. Considering the resources required to conduct such a task and the subjectivity of manual annotation of psychotherapy verbatims, the use of AI algorithms may be an interesting avenue. The main goal to be achieved in this study is to be able to identify the best-performing algorithm to conduct automated annotations of AT verbatims. This requires the proper identification of the best-performing algorithm for the specific context of AT. We hypothesize that support vector machine algorithms will perform best considering the limited dataset available for AT at this time and considering the high number of features being integrated for the automated classification of the interactions taking place in the verbatims.

2. Materials and Methods

2.1. Participants and Recruitment

The data utilized in this study originated from individuals who participated in pilot trials conducted at the Centre de recherche de l'Institut universitaire en santé mentale de Montréal (CR-IUSMM) and an ongoing trial that compares AT to CBT. These participants were enrolled in the clinical trial registered on Clinicaltrials.gov, identified by the number NCT03585127 [15]. All participants received a total of nine one-hour psychotherapeutic sessions, of which eight were immersive sessions involving interaction with a virtual representation of their auditory verbal hallucinations—the Avatar. The participants included in this study were patients of the IUSMM aged over 18 years. They all suffered from treatment-resistant schizophrenia (TRS), defined by the lack of response to two or more dopaminergic antagonists as expressed by the persistence of auditory hallucinations. The AT sessions were administered between the years 2017 and 2022.

2.2. Dataset: Corpus of Avatar Therapy and Features

Immersive sessions of 35 patients who had undergone AT were transcribed verbatim from audio recordings by research auxiliaries. The verbatims were then verified by AH

to ensure the integrity of the transcriptions. This yielded 288 verbatims representing over 250 h of immersion in AT. Annotations of the interactions between the patients and the Avatars were classified as per the 27 themes described in Beaudoin et al. 2021 [20]. The themes are presented in Table 1 for the Avatar and Table 2 for the patients.

Table 1. Summary of Avatar interactions' themes as per Beaudoin et al. 2021 [20].

Avatar Themes	Examples
Accusations	"You did this."
Omnipotence	"I am all over the place."
Beliefs	"I think you are crazy."
Active listening, empathy	"Please relax, take your time."
Incitements, orders	"You should stop doing."
Coping mechanisms	"Tell me why you are sad when I say this?"
Threats	"I will destroy you."
Negative emotions	"It's difficult for me to realize that."
Self-perceptions	"I identify myself as nothing."
Positive emotions	"I am the best in the world".
Provocation	"Try stopping me from making you ill."
Reconciliation	"Should we make peace?"
Reinforcement	"Try this again."

Table 2. Patient interactions' themes as per Beaudoin et al. 2021 [20].

Patient Themes	Examples
Approbation	"You are right."
Self-deprecation	"I can't do this."
Self-appraisal	"I am a nice person."
Other beliefs	"You are the one controlling me."
Counterattack	"You are the one who did this, not me!"
Maliciousness of the voice	"You are trying to make this hard for all."
Negative	"It is very hard."
Negation	"I never did this."
Omnipotence	"I am the greatest."
Disappearance of the voice	"Please leave me alone!"
Positive	"I am feeling wonderful."
Prevention	"I am trying to dismiss you."
Reconciliation of the voice	"Can we work together?"
Self-affirmation	"I am capable of doing this."

A dataset comprising 280 therapy transcripts from thirty-five randomly selected patients who underwent Avatar Therapy (AT) between 2017 and 2022 at our institution was compiled. Each patient participated in eight therapy sessions, resulting in an average of eight transcripts per patient. The transcripts were originally manually typed and were in Canadian French. For annotation purposes, the transcripts were manually annotated using the 27 themes described in the study conducted by Beaudoin et al. in 2021 [20]. The annotation process was carried out using *QDA Miner version 5*, a qualitative data analysis software developed by Provalis Research [24]. The annotations were subsequently

extracted as text files, with each file containing a varying number of interactions (ranging from 1 to 40) related to the same theme. These extracted annotations were then categorized into two conceptual databases: Avatar and Patient, following the representation depicted in Figure 1.

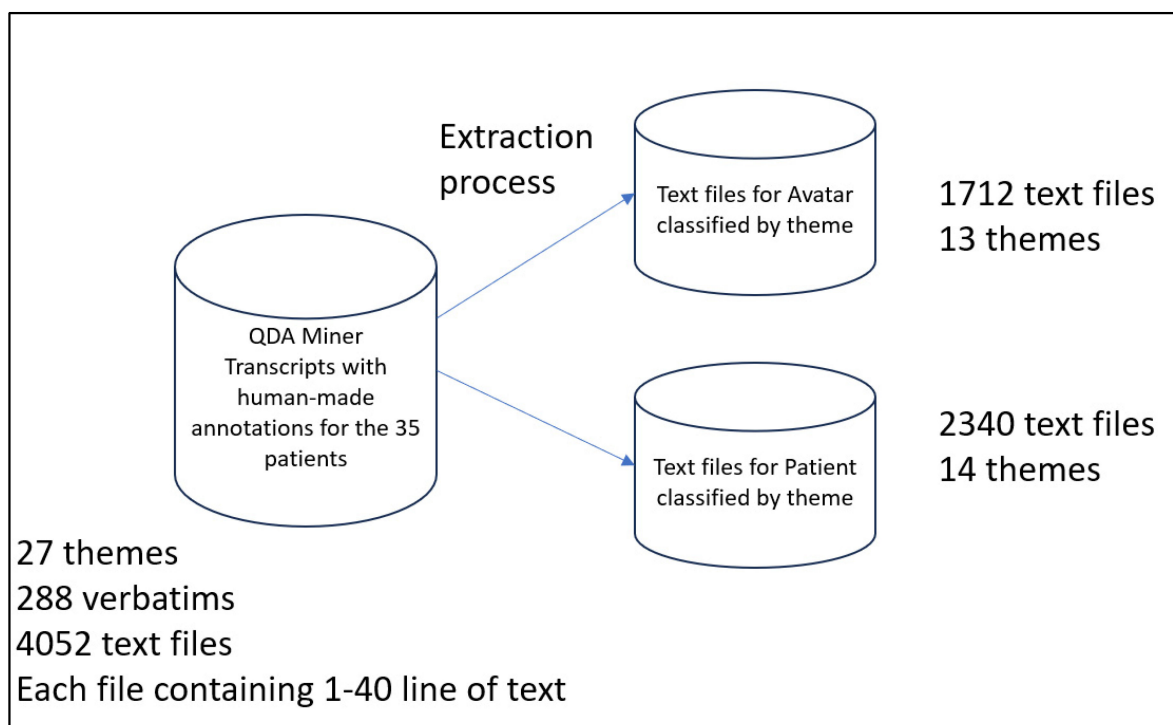


Figure 1. Dataset for the corpus of Avatar Therapy.

2.3. Machine Learning Algorithms

Five algorithms for automated text classification were implemented over the AT dataset in Python 3.11 as per the classification identified in the previous literature review for the context of psychotherapy: Support vector classifier (SVC), Linear support vector classifier (Linear SVC), Multinomial Naïve Bayes (Multinomial NB), Decision Tree (DT), and Multi-layer perceptron classifier (MLP) [23]. They were all used over the Avatar conceptual dataset and the Patient conceptual dataset. A GridSearchCV (GSCV) technique from the Scikit-Learn library was employed to optimize the performance of the machine learning algorithm and improve classification strategies. GSCV is a valuable tool as it allows users to explore various hyperparameters and cross-validate the classifier's predictions, thereby identifying the optimal combination of parameters that yield the best performance. In this study, GSCV was applied to both SVC and LSVC classifiers [25]. Default parameters were utilized for the DT, MLP, and Multinomial NB classifiers, as they demonstrated superior performance when considering hyperparameterization.

The algorithms were paired with a term frequency-inverse document frequency (TF-IDF) statistic, known for its superior performance in text classification when compared with other algorithm-tokenizer combinations. To implement TF-IDF tokenization, we selected the TfidfVectorizer provided by the Scikit-Learn library. This vectorizer facilitates the conversion of the raw text extracted from the interview's interactions into numerical vectors [26]. Additionally, vectorizers can be customized to accommodate stop-words if necessary. Because the classification categories were designed to separate text entities based on their distinct intrinsic characteristics, the assumption is that the features are linearly separable [20].

2.3.1. Support Vector Classifier (SVC)

A Support vector classifier is employed for supervised classification tasks [27]. Finding the best hyperplane to divide several classes of data points in a high-dimensional feature space is the main goal of this particular support vector machine (SVM) approach [28]. Maximizing the margin between classes, it does this with the intention of achieving good generalization performance [29]. It operates by locating a subset of training samples known as support vectors that serve as the decision boundary's key points. These support vectors are critical in choosing the best hyperplane because they are located closest to the decision boundary.

The implementation used for the SVC in this study is from Scikit-Learn, more precisely, the SVC class of the SVM library [26,30].

2.3.2. Linear Support Vector Classifier (Linear SVC)

The Linear support vector classifier belongs to the family of support vector machines. As compared with SVC, Linear SVC uses a linear kernel. A kernel is a mathematical function that is used in a variety of machine-learning methods to turn data into a higher-dimensional feature space [31]. The ability of algorithms to address complicated issues that can be challenging or even impossible to handle in the original input space is fundamentally dependent on kernels. Therefore, a linear kernel is used when the data are linearly separable.

The implementation used for the SVC in this study is from Scikit-Learn, more precisely, the SVC class of the SVM library with the specification of using a linear kernel [30,32].

2.3.3. Multinomial Naïve Bayes Classifier (Multinomial NB)

The main application of the probabilistic machine learning technique known as the Multinomial Naïve Bayes classifier is text classification problems. It is a development of the Naïve Bayes method, which relies on the Bayes theorem and assumes that the characteristics are conditionally independent of the class [33]. The Bayes theorem enables us to revise the likelihood that Event A will occur considering novel data or supporting evidence provided by Event B. By combining the prior probability ($P(A)$) and the likelihood ($P(B|A)$), it offers a method for calculating the posterior probability ($P(A|B)$) [34]. To handle discrete features in text data, such as word counts or frequencies, the Multinomial Naïve Bayes classifier was developed.

The implementation used for the SVC in this study is from Scikit-Learn, more precisely, the MultinomialNB class of the Naïve Bayes library [30].

2.3.4. Decision Tree Classifier (DT)

Decision Tree-based classifiers are non-parametrized and utilized as supervised learning methods for item classification. These classifiers represent observations about an item through branches and draw conclusions about the item's value or score through leaves [35]. The splitting of observations across branches is determined by predefined rules based on the categories used for classification. In the context of text classification, the underlying concept is that each piece of text being classified undergoes a process of splitting across branches until it reaches a leaf (representing a category) according to probabilistic rules established by the designer of the Decision Tree [36].

The implementation used for the DT in this study is from Scikit-Learn, more precisely, the DecisionTreeClassifier class [30].

2.3.5. Multi-Layer Perceptron Classifier (MLP)

A Multi-layer perceptron classifier is used for a variety of machine learning tasks, including classification. It is a model of a feedforward neural network made up of numerous layers of coupled neurons [37]. The input layer, one or more hidden layers, and the output layer are commonly present in the layered structure of the MLP classifier. Multiple neurons make up each layer, which executes calculations on the incoming data and relays the results

to the following layer. Each neuron in each layer of an MLP is connected to every other neuron in the neighboring layers, indicating that the MLP is fully connected. Weights attached to the connections between neurons govern the strength and significance of the information moving through the network [38].

The implementation used for the MLP in this study is from Scikit-Learn, more precisely, the MLPClassifier class from the neural_network library [30].

2.4. Data Analysis and Validation

A partitioning strategy was employed for each conceptual database, where 70% of the annotated documents were used for training the algorithms, while the remaining 30% were utilized for testing purposes [39]. The objective was to establish a statistical probability for each algorithm, represented by a predictive score, indicating the adequacy of classifying an interaction. The training and testing sets were intentionally non-overlapping to adhere to recommended design practices [40,41]. The predictive score corresponds to the average accuracy, measured by the F1-Score, of the themes being evaluated during testing. Additionally, a tenfold cross-validation technique was implemented using the K-Fold model from the Scikit-Learn library for each algorithm [30,42].

The *Classification Report* tool from the Scikit-Learn metrics module was utilized to gather information regarding the classification performance of each theme, including the precision, recall, and F1-Score for each algorithm. Precision represents the positive predictive value, recall indicates the sensitivity of the prediction, and the F1-Score reflects the accuracy of theme classification [43]. The F1-Score is a commonly used measure in text classification that strikes a balance between precision and recall, providing an overall assessment of classification accuracy. The F1-Score is, therefore, the harmonic mean between precision and recall [44].

3. Results

3.1. Sample Characteristics

Interactions taking place in the verbatims of 35 patients were used by the five machine learning algorithms in this study to conduct automated annotation. The characteristics of the sampled patients are found in Table 3.

Table 3. Characteristics of sampled patients.

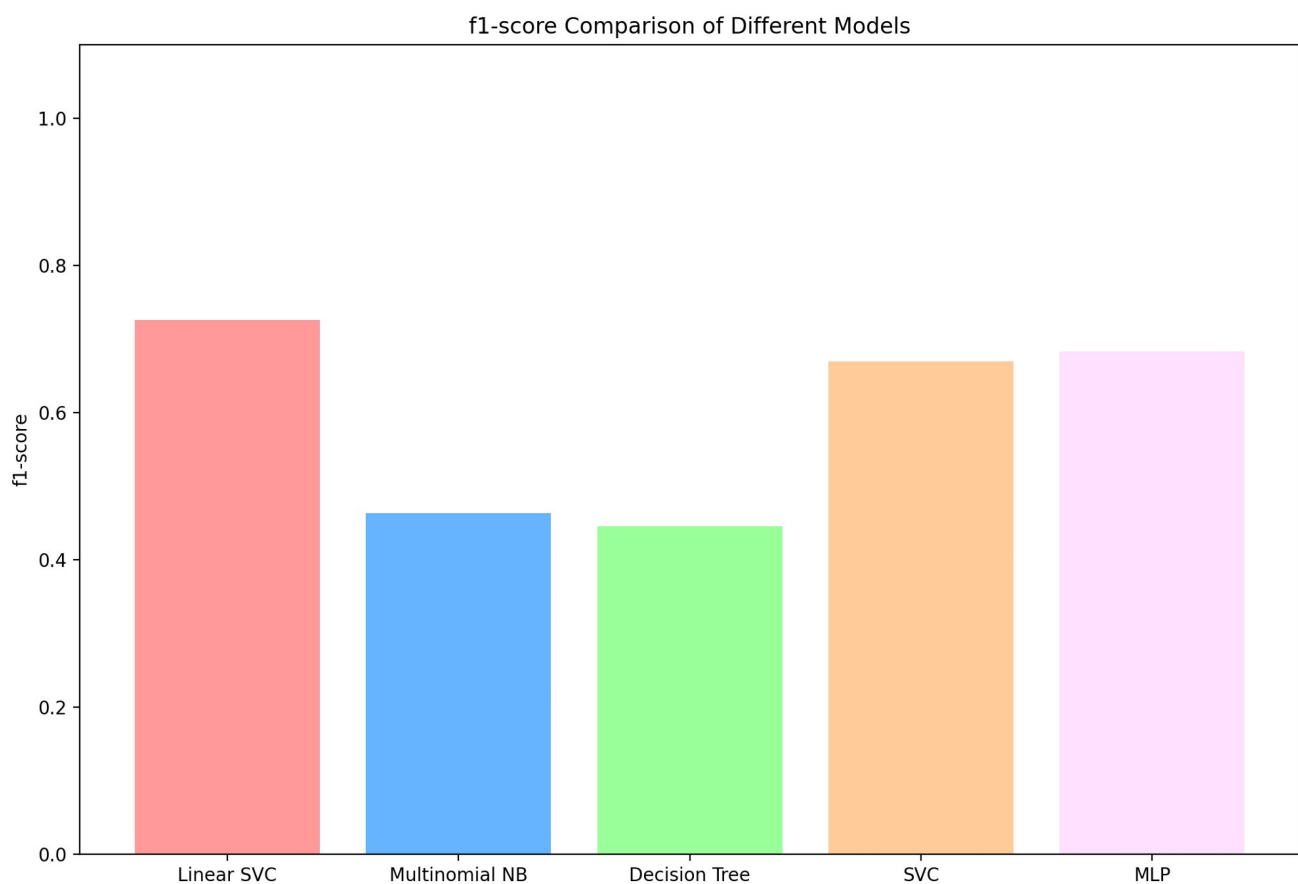
Characteristics	Value ($n = 35$)
Sex (number of males, number of females)	27, 8
Age (mean in years)	41.8 ± 11.2
Education (mean in years)	13.4 ± 3.2
Ethnicity (Caucasian, others)	94.3%, 5.7%
% on clozapine	45.7%

3.2. Performance of Machine Learning Algorithms

The average performance of the machine learning algorithm for the automatic annotation of the verbatim is found in Table 4. It can be observed that the Linear SVC performs best over the dataset as compared with the other algorithms with the highest accuracy score, recall score, and F1-Score. The regular SVC performs best for precision over the dataset. Overall, the DT classifier performs the worst over the analyzed metrics. Descriptive visualization of the F1-Score comparisons can be observed in Figure 2.

Table 4. Average performances of each classifier on the Avatar conceptual database for the metrics: accuracy, precision, recall, and F1-Score.

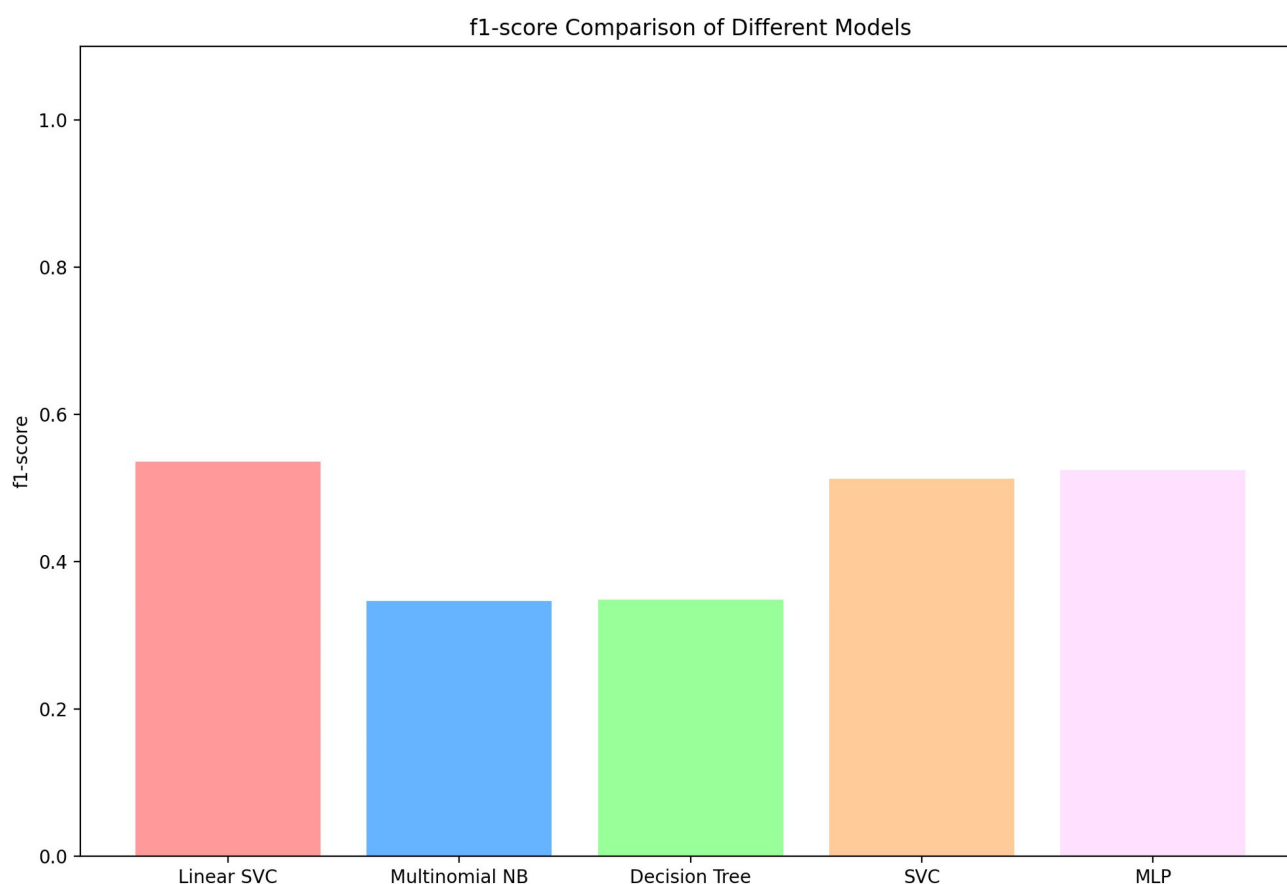
Classifier	Accuracy	Precision	Recall	F1-Score
SVC	0.653680	0.736737	0.636364	0.636396
Linear SVC	0.705628	0.715403	0.675325	0.674928
Multinomial NB	0.437229	0.540432	0.545455	0.488000
Decision Tree	0.350649	0.403547	0.389610	0.388143
MLP	0.662338	0.658041	0.636364	0.636298

**Figure 2.** F1-Score comparisons of the different classifiers over the Avatar database.

The average performances of the different classifiers are presented in Table 5. As for the performance on the Avatar database, it can be observed that the Linear SVC performs best for the F1-Score as well as all the other metrics except for the precision, where the regular SVC offers superior performance. The Decision Tree performs poorly over the database with the smallest F1-Score. Descriptive visualization of the F1-Score comparisons of the models over the Patient dataset can be observed in Figure 3.

Table 5. Average performances of each classifier on the Patient conceptual database for the metrics: accuracy, precision, recall, and F1-Score.

Classifier	Accuracy	Precision	Recall	F1-Score
SVC	0.526842	0.680169	0.526842	0.552448
Linear SVC	0.571930	0.610126	0.571930	0.575930
Multinomial NB	0.315789	0.529961	0.315789	0.297080
Decision Tree	0.350877	0.393063	0.350877	0.359419
MLP	0.564912	0.578114	0.564912	0.567399

**Figure 3.** F1-Score comparisons of the different classifiers over the Patient database.

4. Discussion

This study aimed to compare the performance of machine learning algorithms in the automatic annotation of immersive session verbatims of AT. From the five implementations of machine algorithms over both the Avatar and Patient conceptual databases, it was observed that the Linear SVC performed the best across all metrics except for the precision. The regular SVC performed best for the precision metrics.

Artificial intelligence, especially the field of machine learning, could therefore provide an interesting avenue for automated annotations of psychotherapeutic verbatims, which are usually performed by human coders. This would have the potential to save resources (cost and time) as well as balance subjectivity biases introduced by qualitative assessment of verbatims. Such techniques should be further explored.

While few implementations of supervised machine learning algorithms exist in the clinical applications of psychiatry and psychotherapy, text classification and automated annotation is used in different aspects of medicine. A study by Gibbons et al. (2017) tackled the challenge of classifying open-text feedback of doctor performances with human-

level accuracy on a corpus of 1636 open-text comments relating to the performance of 548 doctors [45]. With a dataset of comparable size as the one used in our study, it was found that their support vector machine classifier (SVM) had a similar F1-Score performance as the one observed in AT. However, in their implementation, DT and the combinations of three and more models yielded better overall performance. This can be explained by the context of their applications of machine learning algorithms' performance comparison, considering they used a context of an open-ended survey as their corpus, which comprised fewer features than the ones used in AT. As complexity grows, algorithms such as SVM-based classifiers perform better in the context of textual entities classified over more features [46,47].

The performance of LSVC over SVC in the context of AT might be intrinsic to the linear separation of the different themes [48]. Considering the previous qualitative analysis conducted on AT, the themes identified were attempted to be as linearly separable as possible. This can explain the overall poor performance of DT and Multinomial NB. A recent review of the application of machine learning algorithms on text classification highlights that Naïve Bayes algorithms often perform poorly, as they assume that all the features are entirely independent of each other, which often is not the case when the corpus is human-generated such as in the context of AT [49]. The Multinomial NB assumes a multinomial distribution of AT interactions that might not be accurate [50]. As for DT, continuous data such as the dataset of this study offers many branching, and this can lead to poor performances. As for the precision performance of SVC over Linear SVC, SVC with an appropriate non-linear kernel can provide better precision by capturing the underlying complexities of the data. The data in AT refers to interactions between the Patient and the Avatar and is intrinsically complex as defined by the underlying naturalistic language being assessed.

Finally, the performance of the MLP might have been impacted by the small size of the database. Neural network algorithm often needs a vast array of data to achieve adequate performance [51].

Limitations

The current analysis of the performance for the different implementations of the machine learning algorithms as described is limited by the small database offered by AT. As more patients are included in the dataset, the trend of the performances for the different algorithms will be re-assessed. It is also important to mention that the transcripts examined in this study were written in Canadian French. A challenge was encountered in finding vectorizers that incorporated stop-words specifically for the Canadian French language. Stop-words are words that are typically excluded from the tokenization process as they hold little or no significant meaning. The absence of appropriate stop-words for Canadian French can potentially impact the accuracy of the analysis, as it may result in insignificant words being included in the word vectors and affecting the overall results.

5. Conclusions

To conclude, this study compared the performances of five machine learning algorithms over the AT dataset. More precisely, it focused on the classification of textual interactions from verbatims of patients suffering from TRS undergoing immersive virtual reality sessions in AT. The Linear SVC algorithm was identified as being the algorithm that performed best in terms of the accuracy, recall, and F1-Score for the Avatar conceptual dataset and the Patient conceptual dataset. The SVC algorithm also performed well compared with the other algorithm, achieving the best performances for precision. This study offers a first comparison of several machine learning algorithms on AT and provides an objective approach to the classification of textual interactions based on immersive session verbatims. Future studies could use this approach to provide insight relating to the elements being classified and the therapeutic response of patients as per their experience with AT immersive sessions.

Author Contributions: Conceptualization, A.H., K.P., S.P. and A.D.; methodology, A.H. and A.D.; validation, A.H. and A.D.; formal analysis, A.H.; investigation, A.H.; data curation, A.H.; writing—original draft preparation, A.H.; writing—review and editing, A.H., K.P., S.P. and A.D.; supervision, K.P., S.P. and A.D.; project administration, K.P.; funding acquisition, K.P., S.P. and A.D. All authors have read and agreed to the published version of the manuscript.

Funding: This work was indirectly supported by Le Fonds de recherche du Québec—Santé (FRQS); Otsuka Canada Pharmaceutical Inc.; Chaire Eli Lilly Canada de recherche en schizophrénie; MEI (Ministère de l'Économie et de l'Innovation); Services et recherches psychiatriques AD; and Fonds d'excellence en recherche Apogée Canada. These funding bodies had no part in the data collection, analysis, interpretation of data, or in writing the manuscript.

Institutional Review Board Statement: This study was approved by the institutional ethical committee, and written informed consent was obtained from all patients. Patients that are part of this study were selected based on the proof-of-concept trial from Percy du Sert's 2018 study and Dellazizzo's 2021 study [15]. The trial was conducted in accordance with the Declaration of Helsinki and was approved by the institutional ethical committee (CER IPPM 16-17-06). We obtained written informed consent from all patients.

Data Availability Statement: The datasets generated and/or analyzed during the current study are not publicly available due to patients' privacy but are available from the corresponding author upon reasonable request.

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

References

- Charlson, F.J.; Ferrari, A.J.; Santomauro, D.F.; Diminic, S.; Stockings, E.; Scott, J.G.; McGrath, J.J.; Whiteford, H.A. Global Epidemiology and Burden of Schizophrenia: Findings from the Global Burden of Disease Study 2016. *Schizophr. Bull.* **2018**, *44*, 1195–1203. [\[CrossRef\]](#)
- Cloutier, M.; Aigbogun, M.S.; Guerin, A.; Nitulescu, R.; Ramanakumar, A.V.; Kamat, S.A.; DeLucia, M.; Duffy, R.; Legacy, S.N.; Henderson, C.; et al. The Economic Burden of Schizophrenia in the United States in 2013. *J. Clin. Psychiatry* **2016**, *77*, 5379. [\[CrossRef\]](#)
- Habtewold, T.D.; Hao, J.; Liemburg, E.J.; Baştürk, N.; Bruggeman, R.; Alizadeh, B.Z. Deep Clinical Phenotyping of Schizophrenia Spectrum Disorders Using Data-Driven Methods: Marching towards Precision Psychiatry. *J. Pers. Med.* **2023**, *13*, 954. [\[CrossRef\]](#) [\[PubMed\]](#)
- Huhn, M.; Nikolakopoulou, A.; Schneider-Thoma, J.; Krause, M.; Samara, M.; Peter, N.; Arndt, T.; Bäckers, L.; Rothe, P.; Cipriani, A.; et al. Comparative efficacy and tolerability of 32 oral antipsychotics for the acute treatment of adults with multi-episode schizophrenia: A systematic review and network meta-analysis. *Lancet* **2019**, *394*, 939–951. [\[CrossRef\]](#) [\[PubMed\]](#)
- Xia, J.; Merinder, L.B.; Belgamwar, M.R. Psychoeducation for schizophrenia. In *Cochrane Database of Systematic Reviews*; John Wiley & Sons, Ltd.: Hoboken, NJ, USA, 2011; p. Cd002831.
- Lally, J.; Gaughran, F.; Timms, P.; Curran, S.R. Treatment-resistant schizophrenia: Current insights on the pharmacogenomics of antipsychotics. *Pharmacogenomics Pers. Med.* **2016**, *9*, 117–129. [\[CrossRef\]](#)
- Potkin, S.G.; Kane, J.M.; Correll, C.U.; Lindenmayer, J.P.; Agid, O.; Marder, S.R.; Olfson, M.; Howes, O.D. The neurobiology of treatment-resistant schizophrenia: Paths to antipsychotic resistance and a roadmap for future research. *NPJ Schizophr.* **2020**, *6*, 1. [\[CrossRef\]](#) [\[PubMed\]](#)
- Stepnicki, P.; Kondej, M.; Kaczor, A.A. Current Concepts and Treatments of Schizophrenia. *Molecules* **2018**, *23*, 2087. [\[CrossRef\]](#)
- Guaiana, G.; Abbatecola, M.; Aali, G.; Tarantino, F.; Ebuanyi, I.D.; Lucarini, V.; Li, W.; Zhang, C.; Pinto, A. Cognitive behavioural therapy (group) for schizophrenia. *Cochrane Database Syst. Rev.* **2022**, *7*, Cd009608.
- Aali, G.; Kariotis, T.; Shokraneh, F. Avatar Therapy for people with schizophrenia or related disorders. *Cochrane Database Syst. Rev.* **2020**, *5*, Cd011898.
- Dellazizzo, L.; Potvin, S.; Phraxayavong, K.; Lalonde, P.; Dumais, A. Avatar Therapy for Persistent Auditory Verbal Hallucinations in an Ultra-Resistant Schizophrenia Patient: A Case Report. *Front. Psychiatry* **2018**, *9*, 131. [\[CrossRef\]](#)
- Leff, J.; Williams, G.; Huckvale, M.; Arbuthnot, M.; Leff, A.P. Avatar therapy for persecutory auditory hallucinations: What is it and how does it work? *Psychosis* **2014**, *6*, 166–176. [\[PubMed\]](#)
- Leff, J.; Williams, G.; Huckvale, M.A.; Arbuthnot, M.; Leff, A.P. Computer-assisted therapy for medication-resistant auditory hallucinations: Proof-of-concept study. *Br. J. Psychiatry* **2013**, *202*, 428–433. [\[CrossRef\]](#) [\[PubMed\]](#)
- Craig, T.K.; Rus-Calafell, M.; Ward, T.; Leff, J.P.; Huckvale, M.; Howarth, E.; Emsley, R.; Garety, P.A. AVATAR therapy for auditory verbal hallucinations in people with psychosis: A single-blind, randomised controlled trial. *Lancet Psychiatry* **2018**, *5*, 31–40. [\[CrossRef\]](#)
- Dellazizzo, L.; Potvin, S.; Phraxayavong, K.; Dumais, A. One-year randomized trial comparing virtual reality-assisted therapy to cognitive-behavioral therapy for patients with treatment-resistant schizophrenia. *NPJ Schizophr.* **2021**, *7*, 9. [\[CrossRef\]](#)

16. Chai, H.H.; Gao, S.S.; Chen, K.J.; Duangthip, D.; Lo, E.C.M.; Chu, C.H. A Concise Review on Qualitative Research in Dentistry. *Int. J. Environ. Res. Public Health* **2021**, *18*, 942. [\[CrossRef\]](#)
17. Pannucci, C.J.; Wilkins, E.G. Identifying and Avoiding Bias in Research. *Plast Reconstr. Surg.* **2010**, *126*, 619–625. [\[CrossRef\]](#)
18. Starks, H.; Trinidad, S.B. Choose your method: A comparison of phenomenology, discourse analysis, and grounded theory. *Qual. Health Res.* **2007**, *17*, 1372–1380. [\[CrossRef\]](#) [\[PubMed\]](#)
19. Dellazizzo, L.; Percie du Sert, O.; Phraxayavong, K.; Potvin, S.; O'Connor, K.; Dumais, A. Exploration of the dialogue components in Avatar Therapy for schizophrenia patients with refractory auditory hallucinations: A content analysis. *Clin. Psychol. Psychother.* **2018**, *25*, 878–885. [\[CrossRef\]](#)
20. Beaudoin, M.; Potvin, S.; Machalani, A.; Dellazizzo, L.; Bourguignon, L.; Phraxayavong, K.; Dumais, A. The therapeutic processes of avatar therapy: A content analysis of the dialogue between treatment-resistant patients with schizophrenia and their avatar. *Clin. Psychol. Psychother.* **2021**, *28*, 500–518. [\[CrossRef\]](#) [\[PubMed\]](#)
21. Sidey-Gibbons, J.A.; Sidey-Gibbons, C.J. Machine learning in medicine: A practical introduction. *BMC Med. Res. Methodol.* **2019**, *19*, 64. [\[CrossRef\]](#) [\[PubMed\]](#)
22. Chekroud, A.M.; Bondar, J.; Delgadillo, J.; Doherty, G.; Wasil, A.; Fokkema, M.; Cohen, Z.; Belgrave, D.; DeRubeis, R.; Iniesta, R.; et al. The promise of machine learning in predicting treatment outcomes in psychiatry. *World Psychiatry* **2021**, *20*, 154–170. [\[CrossRef\]](#)
23. Hudon, A.; Beaudoin, M.; Phraxayavong, K.; Dellazizzo, L.; Potvin, S.; Dumais, A. Use of Automated Thematic Annotations for Small Data Sets in a Psychotherapeutic Context: Systematic Review of Machine Learning Algorithms. *JMIR Ment. Health* **2021**, *8*, e22651. [\[CrossRef\]](#) [\[PubMed\]](#)
24. Lewis, R.B.; Maas, S.M. QDA Miner 2.0: Mixed-model qualitative data analysis software. *Field Methods* **2007**, *19*, 87–108. [\[CrossRef\]](#)
25. Paper, D.; Paper, D. Scikit-Learn Classifier Tuning from Simple Training Sets. In *Hands-on Scikit-Learn for Machine Learning Applications: Data Science Fundamentals with Python*; Apress: Berkeley, CA, USA, 2020; pp. 137–163.
26. Hutter, F.; Kotthoff, L.; Vanschoren, J. Hyperopt-sklearn. In *Automated Machine Learning: Methods, Systems, Challenges*; Springer: Berlin/Heidelberg, Germany, 2019; p. 219.
27. Mammone, A.; Turchi, M.; Cristianini, N. Support vector machines. *Wiley Interdiscip. Rev. Comput. Stat.* **2009**, *1*, 283–289. [\[CrossRef\]](#)
28. Shao, Y.H.; Chen, W.J.; Deng, N.Y. Nonparallel hyperplane support vector machine for binary classification problems. *Inf. Sci.* **2014**, *263*, 22–35. [\[CrossRef\]](#)
29. Xu, J.; Liu, X.; Huo, Z.; Deng, C.; Nie, F.; Huang, H. Multi-class support vector machine via maximizing multi-class margins. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence (IJCAI 2017)*, Melbourne, Australia, 19–25 August 2017; pp. 3154–3160.
30. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
31. Almaiah, M.A.; Almomani, O.; Alsaaidah, A.; Al-Otaibi, S.; Bani-Hani, N.; Hwaitat, A.K.A.; Al-Zahrani, A.; Lutfi, A.; Awad, A.B.; Aldhyani, T.H. Performance investigation of principal component analysis for intrusion detection system using different support vector machine kernels. *Electronics* **2022**, *11*, 3571. [\[CrossRef\]](#)
32. Varoquaux, G.; Buitinck, L.; Louppe, G.; Grisel, O.; Pedregosa, F.; Mueller, A. Scikit-learn: Machine learning without learning the machinery. In *GetMobile: Mobile Computing and Communications*; Association for Computing Machinery: New York, NY, USA, 2015; Volume 19, pp. 29–33.
33. Rish, I. An empirical study of the naive Bayes classifier. In *Proceedings of the IJCAI 2001 Workshop on Empirical Methods in Artificial Intelligence*, Seattle, WA, USA, 4 August 2001.
34. Berrar, D. Bayes' theorem and naive Bayes classifier. In *Encyclopedia of Bioinformatics and Computational Biology*; Elsevier: Amsterdam, The Netherlands, 2019; pp. 403–412.
35. Kingsford, C.; Salzberg, S.L. What are decision trees? *Nat. Biotechnol.* **2008**, *26*, 1011–1013. [\[CrossRef\]](#)
36. Kotsiantis, S.B. Decision trees: A recent overview. *Artif. Intell. Rev.* **2013**, *39*, 261–283. [\[CrossRef\]](#)
37. Ramchoun, H.; Ghanou, Y.; Ettaouil, M.; Janati Idrissi, M.A. Multilayer perceptron: Architecture optimization and training. *Int. J. Interact. Multimed. Artif. Intell.* **2016**, *4*, 26–30. [\[CrossRef\]](#)
38. Popescu, M.C.; Balas, V.E.; Perescu-Popescu, L.; Mastorakis, N. Multilayer perceptron and neural networks. *WSEAS Trans. Circuits Syst.* **2009**, *8*, 579–588.
39. Gholamy, A.; Kreinovich, V.; Kosheleva, O. Why 70/30 or 80/20 relation between training and testing sets: A pedagogical explanation. *Int. J. Inf. Technol. Appl. Sci.* **2018**, *11*, 1–6.
40. Bhavsar, H.; Ganatra, A. A comparative study of training algorithms for supervised machine learning. *Int. J. Soft Comput. Eng. (IJSCE)* **2012**, *2*, 2231–2307.
41. Huang, X.; Jin, G.; Ruan, W. Machine Learning Basics. In *Machine Learning Safety*; Artificial Intelligence: Foundations, Theory, and Algorithms Book Series; Springer: Singapore, 2012; pp. 3–13.
42. Wong, T.T.; Yeh, P.Y. Reliable accuracy estimates from k-fold cross validation. *IEEE Trans. Knowl. Data Eng.* **2019**, *32*, 1586–1594. [\[CrossRef\]](#)
43. Goutte, C.; Gaussier, E. A probabilistic interpretation of precision, recall and F-score, with implication for evaluation. In *European Conference on Information Retrieval*; Springer: Berlin/Heidelberg, Germany, 2005; pp. 345–359.

44. Opitz, J.; Burst, S. Macro f1 and macro f1. *arXiv* **2019**, arXiv:1911.03347.
45. Gibbons, C.; Richards, S.; Valderas, J.M.; Campbell, J. Supervised Machine Learning Algorithms Can Classify Open-Text Feedback of Doctor Performance with Human-Level Accuracy. *J. Med. Internet Res.* **2017**, *19*, e6533. [[CrossRef](#)]
46. Joachims, T. Text categorization with support vector machines: Learning with many relevant features. In *Machine Learning: ECML-98; Lecture Notes in Computer Science Book Series*; Springer: Berlin/Heidelberg, Germany, 1998; Volume 1389, pp. 137–142.
47. Liu, Z.; Lv, X.; Liu, K.; Shi, S. Study on SVM compared with the other text classification methods. In Proceedings of the 2010 Second International Workshop on Education Technology and Computer Science, Wuhan, China, 6–7 March 2010; pp. 219–222.
48. Amarappa, S.; Sathyanarayana, S.V. Data classification using Support vector Machine (SVM), a simplified approach. *Int. J. Electron. Comput. Sci. Eng.* **2014**, *3*, 435–445.
49. Li, R. *A Review of Machine Learning Algorithms for Text Classification*; Springer Nature: Singapore, 2022.
50. Harzevili, N.S.; Alizadeh, S.H. Mixture of latent multinomial naive Bayes classifier. *Appl. Soft Comput.* **2018**, *69*, 516–527. [[CrossRef](#)]
51. Singh, Y.; Chauhan, A.S. Neural Networks in Data Mining. *J. Theor. Appl. Inf. Technol.* **2009**, *5*, 36–42. [[CrossRef](#)]

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