Development and Deployment of Sentiment Analysis AI on Citizens' Feedback in Goiás

Rafaela M. Rosa [a], *Jessé D. de* Souza [b], *Camila do N.* Freitas [c], etc.

[a] State Government of Goiás, Secretariat of Administration, Goiás, Brazil, rafaela.rosa@goias.gov.br, ORCID number: 0009-0007-9040-4189.

[b] State Government of Goiás, Secretariat of Administration, Goiás, Brazil, jesse.souza@goias.gov.br, ORCID number: 0009-0007-3802-6117.

[c] State Government of Goiás, Secretariat of Administration, Goiás, Brazil, camila.freitas@goias.gov.br, ORCID number: 0009-0004-2208-7167.

Submitted: 31 January 2025, Revised: 26 March 2025, Accepted: 21 April 2025, Published: 19 May 2025

Abstract. This study presents the development and deployment of a sentiment analysis model based on Natural Language Processing (NLP) techniques in the context of public service evaluation. The model was implemented to automate the classification of citizens' feedback on services provided by the EXPRESSO program in the State of Goiás, Brazil. Before implementation, the classification process was manual, which was time-consuming, and prone to inconsistencies. The dataset used included a wide range of citizen experiences, ensuring that the model captured a comprehensive and representative sample of comments. Preprocessing techniques such as noise removal, tokenization, stemming, and customized stopword adjustments were applied to refine the data for analysis. Among the algorithms tested, Multinomial Naïve Bayes (MNB) stood out for its slightly superior performance in identifying negative feedback — a critical class for monitoring service quality and addressing citizen concerns efficiently. The model was validated with managers to ensure its practical application and was subsequently integrated into the State's Data Warehouse (DW) to feed a real-time monitoring dashboard. Additionally, a visual interface was developed to allow managers to analyze feedback on demand. This interface includes features such as word clouds for positive and negative feedback, facilitating the quick identification of key themes, trends, and recurring issues. The integration of artificial intelligence into the EXPRESSO Program establishes a scalable and replicable framework to improve decision-making processes, enhance operational efficiency, and promote inclusivity. This methodology can be extended to other public service centers, emphasizing the vital role of AI in fostering responsive governance, improving service quality, and enhancing citizens' satisfaction.

Keywords. Sentiment analysis, Natural Language Processing, public service evaluation, machine learning, Naive Bayes.
Research paper, DOI: https://doi.org/10.59490/dgo.2025.944

1. Introduction

The use of advanced technologies has redesigned the way governments deal with challenges in the delivery of public services. In this scenario, Artificial Intelligence (AI) has proven to be fundamental for processing large volumes of data and identifying relevant patterns, becoming a strategic tool in decision-making. The application of AI in public administration has contributed to critical areas, such as health, urban mobility and with the analysis of sentiments, citizens' feedback and real demands can now be monitored, promoting greater accessibility and inclusion, as well as better services to the public.

Copyright ©2025 by the authors. This conference paper is published under a CC-BY-4.0 license

This paper presents the development and deployment of a sentiment analysis model based on Natural Language Processing (NLP) in the EXPRESSO Program, of the Government of the State of Goiás. The initiative used a representative sample of feedback from service channels spread throughout the State, reflecting the diversity of citizens' demands. The implemented solution automates the classification of comments into three categories — positive, neutral, and negative — allowing managers to quickly identify critical points in the services offered and proactively meet the main demands.

The approach adopted sought to optimize processes and also promote more inclusive public management, by giving voice to citizens and ensuring that their experiences and opinions are taken into account in the improvement of services. Before the implementation of the model, the classification of these comments was carried out manually, being time-consuming and being subject to inconsistencies. Now, with AI, it is possible to process large volumes of data in seconds, increasing the accuracy and efficiency of the analysis. The methodology developed proved to be scalable and replicable, with the potential of expansion to other call centers, which is the most widely adopted public service model throughout Brazil.

This work reinforces the role of AI in promoting a more efficient public administration that is guided by the real demands of the population. It also evidences how technology can be a powerful tool to give visibility to citizens' voices, making public services more transparent, accessible, and inclusive.

2. Related Work

2.1 Artificial Intelligence

Artificial Intelligence (AI) is a field of computer science focused on developing systems capable of performing tasks that typically require human intelligence. These tasks include learning, pattern recognition, decision-making, and problem-solving. AI leverages advanced algorithms and large volumes of data to train models that simulate aspects of human reasoning (Russell & Norvig, 2021).

The integration of artificial intelligence (AI) into public services is increasingly recognized as a solution to enhance interactions between citizens and the government. O'Neill (2021) highlights that AI adoption in public administrations can optimize processes while improving the government's ability to meet the population's needs more effectively. AI-driven systems, such as chatbots and predictive data analysis, can streamline public service operations, enhancing the quality of services provided.

Chien and Chen (2020) demonstrate that digital transformation in the public sector, driven by AI, is not limited just to task automation, but it involves the creation of new governance models. AI-based approaches have emerged as crucial monitoring tools, enabling public managers to identify patterns and societal needs with greater precision. This transformation fosters greater transparency in public administration and enhances communication efficiency alongside the society.

These innovative approaches allow public managers to better understand societal perception, identify priority areas for improvement, and develop targeted strategies to enhance public service delivery. Such initiatives reinforce the importance of leveraging technological tools to create a public administration that is more responsive to citizens' needs and fosters a better government-society engagement.

Models that Inspired the Use of AI in Public Services:

- **United Kingdom – National Health Service (NHS):** The UK's National Health Service employs AI to improve medical diagnosis accuracy and forecast future hospital demands. The "Streams" system, developed in collaboration with DeepMind, analyzes patient data in real-time to identify signs of acute kidney failure, enabling quicker interventions. (Royal Free London NHS Foundation Trust & DeepMind. "Our work on Streams." 2021. Available at: https://www.nhs.uk)
- **United States – Internal Revenue Service (IRS):** In the United States, the IRS has adopted AI systems to detect tax fraud and enhance customer service. By employing advanced algorithms, the IRS identifies irregularities in tax returns more efficiently, saving millions of dollars in manual audits. (Internal Revenue Service. "IRS Artificial Intelligence in Tax Compliance." 2022. Available at: https://www.irs.gov)
- **Estonia – X-Road System:** Estonia, a leader in digital transformation, utilizes AI in its X-Road system to integrate public services. Through this platform, citizens can access a wide range of government services online, from business registration to medical consultations. AI analyses service usage to optimize access and identify areas for expansion. (E-Estonia. "X-Road: The Backbone of Digital Estonia." 2022. Available at: https://e-estonia.com)

Copyright ©2025 by the authors. This conference paper is published under a CC-BY-4.0 license

- **India – Ayushman Bharat Health System:** In India, the Ayushman Bharat program, focused on healthcare, uses AI to monitor and predict disease outbreaks and optimize medical resource planning in underserved regions. This initiative has saved lives by anticipating public health crises. (National Health Authority of India. "Leveraging AI in Ayushman Bharat." 2022. Available at: https://www.pmjay.gov.in)
- **China – Urban Monitoring and Intelligent Mobility:** Several cities in China, such as Hangzhou, employ AI systems for intelligent mobility, including Alibaba's "City Brain" platform. This system analyses real-time traffic data to reduce congestion and improve emergency response. As a result, travel times in urban areas have significantly decreased. (Alibaba Cloud. "City Brain: Smart Cities Powered by AI." 2021. Available at: https://www.alibabacloud.com)

2.2 Machine Learning

Machine Learning is a subfield of artificial intelligence that enables computational systems to learn and improve their performance when exposed to data and past experiences (Alpaydin, E. (2020). Introduction to Machine Learning. MIT Press). This technology is widely used to solve complex problems, automate processes, and perform tasks involving the transformation or classification of large volumes of data.

Machine learning provides the foundation for systems to learn from data and enhance their performance as they process new information. When applied to the field of **Natural Language Processing (NLP)**, this capability is expanded, allowing computers to understand and analyze human text automatically. One of the possible applications of NLP in public administration is **Sentiment Analysis**, which can interpret emotions and sentiments expressed in the citizens' comments. This becomes a valuable tool for assisting public managers in understanding citizens' demands for improved public services.

2.3 Sentiment Analysis

Sentiment Analysis, according to Verma (2022), plays an important role in promoting innovation, transparency, greater citizen participation, and efficiency in public administration. This technology not only enhances communication between the government and society but also provides practical solutions to complex problems, such as traffic congestion, crime prediction, and disaster management. Applications in areas such as **smart governance, smart mobility, smart infrastructure, and smart living** serve as the foundation for the development of more connected cities.

Verma (2022) also highlights that by combining **supervised learning** techniques, sentiment analysis enables the transformation of large volumes of data into actionable insights. Additionally, it can assist in developing strategies to create smarter and more inclusive environments and services.

2.4 Multinomial Naïve Bayes Classifier

The Multinomial Naïve Bayes (MNB) model is based on Bayes' Theorem for classification tasks. Known for its simplicity and efficiency, it is referred to as "Naïve" due to its assumption of independence between features. The algorithm estimates the probability of assigning a given item to a specific class, particularly excelling in text classification based on the words contained in documents.

The multinomial approach stands out by considering the frequency of word occurrences in a document rather than treating words merely as present or absent. This characteristic makes it particularly effective in contexts where term repetition conveys meaningful information for categorization (ABBAS et al., 2019). In addition to this feature, MNB is easy to interpret and demonstrates good processing performance, making it a widely used tool in text classification applications (ARAR & AYAN, 2017). Notably, MNB achieved the highest accuracy in categorizing "negative" feedback, which was the target class for monitoring public services provided by the government of Goiás.

Tab. 1 - Comments and Occurrences by Polarity.

Polarity	Comments	Total Occurrences
Positive	"O atendente foi muito educado." (Attendant was very polite) / "O serviço foi rapido." (The service was fast)	2
Negative	"O sistema é lento e continua travando." (System is slow and keeps getting stuck) / "Terrível." (Terrible)	3
Neutral	"Queda de energia afetou computadores." (Power outage affected computers)	1

Copyright ©2025 by the authors. This conference paper is published under a CC-BY-4.0 license

Probabilities:

- Positive: P (pos) = 2/6 = 0,33P
- Negative: P(neg) = 3/6 = 0,50
- Neutral: P(neu) = 1/6 ≈0,17

MNB utilizes message frequencies to calculate the probabilities of each polarity. This method serves as the foundation for estimating conditional probabilities in text classification.

3. EXPRESSO: Citizen Service Program and Its Evaluation Methodology

In October 1999, the first **Vapt Vupt** unit was inaugurated at Buriti Shopping, located in Aparecida de Goiânia, Goiás, with the primary objective of consolidating and simplifying access to public services at different levels—federal, state, and municipal—while also including private companies that serve public interests. Since then, the integrated citizen service model developed in the State of Goiás has become a benchmark in the public services delivery. (https://vaptvupt.go.gov.br/sobre-vaptvupt)

In Brazil, integrated service center models have become a reference in public service delivery, with the State of Goiás being one of the pioneers in adopting this approach through **Vapt Vupt**. According to Lara and Gosling (2016), the management of the relationship between citizens and public administration should be structured based on strategies that promote greater citizen participation and satisfaction, aligning services with their expectations. The integrated service system created in Goiás has counterparts in other brazilian states, such as **Poupatempo** in São Paulo, the **Serviço de Atendimento ao Cidadão** (SAC) in Bahia, the **Organização de Centrais de Atendimento** (OCA) in Acre, the **Unidades de Atendimento Integrado** (UAI) in Minas Gerais, **Ganha Tempo** in Mato Grosso, **Ceará Cidadão** in Ceará, and **Na Hora** in the Federal District.

Goiás State Law No. 20.846, enacted on September 2, 2020 (Goiás, 2020), established the citizen service policy in the State of Goiás and created the **EXPRESSO Program**. This program introduced a new model of service delivery to citizens, integrating civil servants, services, and users in a fast, practical, and decentralized manner. Currently, six service channels are available: **Vapt Vupt, Balcão (Service Desks), Correios (Post Office), Web, APP**, and **Self-Service Totens**.

Each of these channels offers a different service approach, with some being in-person (**Vapt Vupt, Balcão, and Correios**) while others provide digital services (**Web, App, and Self-Service Totens**). Each channel serves a specific audience: in-person channels cater to citizens in the cities where they are located as well as in nearby regions, whereas digital channels are accessible to all residents of the State of Goiás.

At the end of the service, citizens are given the option to evaluate their experience and provide feedback. The requirements and evaluation criteria for EXPRESSO services were established by the **Secretariat of Administration (SEAD)**, in compliance with State Law's Article 59, which considers the following aspects:

- **User satisfaction**
- **Service quality**
- **Compliance with commitments and deadlines established for execution**
- **Number of user interactions**
- **Measures adopted by the public administration for service improvement and enhancement**

The evaluation methodology employs the **NPS (Net Promoter Score) scale**, developed by Reichheld (2003), a well-established tool for measuring user satisfaction. First, the user assigns a rating, followed by additional questions to understand the reasons behind the given score. The results classify users into three distinct groups:

- **Promoters (score of 9 or 10):** Users who positively evaluate the service and often provide constructive suggestions.
- **Neutrals (score of 7 or 8):** Users who evaluate the service but typically do not make recommendations or do so with reservations.
- **Detractors (score of 0 to 6):** Users who are dissatisfied with the service

According to **SEAD – State Secretariat of Administration (2024),** the evaluation form presented to citizens is structured as follows:

Copyright ©2025 by the authors. This conference paper is published under a CC-BY-4.0 license

- **NPS question (0 to 10) to assess the premises physical infrastructure.** After selecting a score, follow-up questions are provided to justify the rating.
- **NPS question (0 to 10) to assess service delivery.** Additional questions follow to justify the given score.
- At this point, the citizen is invited to evaluate the service and/or leave a comment, both optional.
- If the citizen chooses to evaluate the service, an **NPS question (0 to 10)** is presented, followed by justification questions.
- At the end of the form, the citizen is once again invited to leave a comment, which is optional

Since most of the questions are quantitative, they are analyzed using statistical methods. Comments were manually reviewed by the technical team, which applied manual filters to remove noise, blank comments, or those containing irrelevant words. After filtering, the team individually analyzed each comment, classifying them into the following categories: **Praise, Inquiry, Complaint, Suggestion**, or a combination such as **Praise/Complaint** and **Suggestion/Complaint**. These last two classifications were used when a citizen addressed two issues simultaneously within the same comment.

From May 2022 to December 2023, a total of **71,785,073 services** were provided through the EXPRESSO channels. Since feedback is optional, **437,627 feedback responses** were received during this period, representing **0.61% of the total services provided**. Within this timeframe, **95,000 comments** were submitted, accounting for **0.13% of services and 21.70% of the total feedback responses**.

Although proportionally small, the sample is statistically significant. When considering only the feedback data **(437,627 out of 71,785,073 services)**, this citizen survey demonstrates a **95% confidence level** with a **margin of error of just 0.15%**. When analyzing only the data collected through comments, the margin of error increases to **0.32 percentage points**, but the confidence level remains at **95%**. Given that the feedback originates from both digital and in-person service channels across multiple locations in the State, it is reasonable to assert that the collected feedback is representative.

Monthly reports on user feedback regarding EXPRESSO services are published via https://www.go.gov.br/ and https://goias.gov.br/administracao/transformacao-da-gestao-publicos, from which the numbers used in this research were obtained.

Based on these findings, the **Subsecretariat for Innovation in Management and Public Services** (In portuguese Subsecretaria de Inovação da Gestão e dos Serviço Público) of SEAD identified an opportunity to automate the classification of citizen comments on EXPRESSO services, optimizing the manual process. Before implementing the new methodology, this process took an average of **two working days per week for each monthly batch of comments**. The manual approach was time-consuming and prone to inconsistencies due to human subjectivity, especially in ambiguous classifications such as sentiment combinations.

Given this scenario, the primary objective was to **apply AI techniques to automate comment classification**, reducing processing time and increasing result consistency. A project was then proposed to develop a sentiment **analysis model**, named **ANE – Análise Expresso** (EXPRESSO Analysis), leveraging **NLP** concepts. This approach enabled the reclassification of comments into three **more objective** categories: **Positive, Neutral, and Negative**.

4. Research Methodology

The actions carried out in this research can be summarized in the following processes: data selection, data preprocessing and cleaning, data transformation, training and evaluation of classification algorithms, classification validation with managers, and finally, script adaptation for production.

Understanding the problem and the dataset is essential for applying NLP techniques. An initial exploratory analysis was conducted to identify and address data inconsistencies, textual noise, and critical issues such as class imbalance.

The development of the sentiment analysis model was completed in approximately eight months and employed supervised learning, as it is widely used in sentiment classification problems and is highly efficient when working with labeled data (LIU, 2012). With supervised learning, manually classified examples are provided during training, allowing parameter adjustments to refine the mapping of textual features to the target classes. This approach also ensures scalability, enabling the application of the model to large volumes of future data.

4.1 Data selection

The analyzed comments were collected directly from the EXPRESSO feedback systems in .csv format, with our main

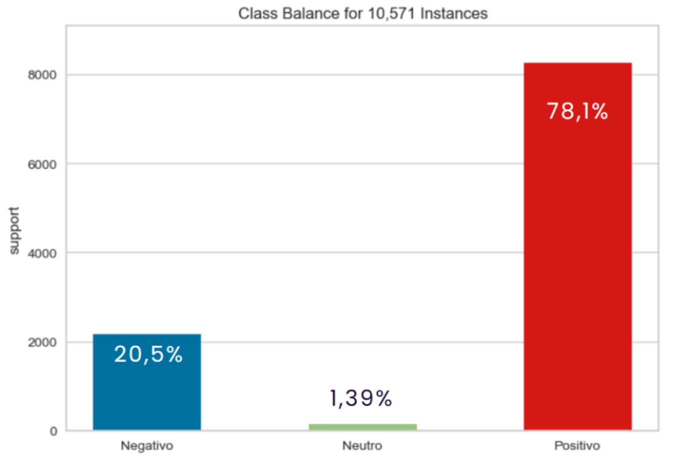
Copyright ©2025 by the authors. This conference paper is published under a CC-BY-4.0 license

focus on two columns: "Service Channel" and "Leave your comment here", the latter being our primary column containing citizen's feedback. For the initial study, a two-month sample of comments—June and July 2023—was used, totaling 10,571 comments.

Initially, a task force was organized to adjust the targets of the previous manual classification, which included six initial categories (PRAISE, QUESTION, COMPLAINT, SUGGESTION, or combinations such as PRAISE/COMPLAINT and SUGGESTION/COMPLAINT). This step was crucial for methodological approaches in Supervised Learning, allowing the model to be trained with greater accuracy. The comments were then reviewed and reclassified into three categories: POSITIVE, NEUTRAL, and NEGATIVE. This new classification became the updated target, defined as **Previsão_de_Sentimentos**, now containing only three classes.

During the initial analysis, we identified a significant imbalance in the dataset, as shown in Fig. 1. Most of the received comments had a positive tone, while the examples classified as negative—our primary focus for identifying opportunities to improve public services—were underrepresented.

Fig. 1 - Distribution of comments (June/July 2023).



This imbalance interfered with the model's training, leading to overfitting issues. In other words, the model performed well on training data but showed lower performance on new data, indicating poor generalization. This behavior often occurs in imbalanced or insufficiently representative datasets (GOODFELLOW et al., 2016).

To address this issue, we opted to strategically expand and refine the dataset. We requested additional feedback from the EXPRESSO system, incorporating comments from 03/2022 to 12/2023—totaling approximately 95,000 comments—with the goal of obtaining a higher number of negative examples. This strategy resulted in a more representative and balanced dataset.

4.2 Preprocessing

Textual databases often present challenges related to the quality and structure of texts. Issues such as grammatical errors, duplications, and the presence of irrelevant terms are common, especially in citizen comments that use informal language.

To address these issues, we applied a sequence of specific treatments in the preprocessing phase, which involved the following activities:

a) Noise removal: Irrelevant elements for analysis were removed, such as special characters, isolated numbers, excessive repetitions of letters or words (e.g. "sooo good!"), URLs, and other symbols that could introduce noise into the data. This step was essential to clean up the text and ensure greater consistency in the data (AGGARWAL & ZHAI, 2012).

b) Text normalization: The texts were converted to lowercase, unnecessary punctuation was removed, and linguistic standardizations were applied. This step helped to standardize the data format, reducing variability generated by different writing styles and typographical errors.

c) Tokenization: The texts were then divided into smaller units, called "tokens," which can represent words, phrases, or even characters, depending on the level of detail required. For this process, we used the Natural Language Toolkit library (NLTK), a specific module for tokenization, called tokenize. The NLTK library is widely used due to its broad application in NLP. Below is a demonstration of the tokenization process:

Copyright ©2025 by the authors. This conference paper is published under a CC-BY-4.0 license

Original text: Ótimo atendimento, foi tudo rápido e prático! (Great service, everything was fast and practical!)
Punctuation removal: Ótimo atendimento foi tudo rápido e prático
Tokenization into individual words: [Ótimo, atendimento, foi, tudo, rápido, e, prático]
Resulting tokens: ["Ótimo", "atendimento", "foi", "tudo", "rápido", "e", "prático"]

d) **Stopwords:** Stopwords are common words, such as articles, prepositions, and conjunctions, that generally have little informational value (Pang & Lee, 2008). When processing large volumes of text, removing stopwords helps with the analysis and also saves computational resources. However, this step needs to be carefully considered based on the objective, as some words, like "não" (not), can completely alter the meaning of a sentence in sentiment analysis. By removing the word 'não', phrases like 'não é bom' (it's not good) or 'não gostei do atendimento' (I did not like the service) would lose their negative context, hindering the model's ability to correctly identify the sentiment polarity. To solve this problem, we adjusted the default stopwords list. This approach improves the effectiveness of sentiment analysis, particularly in scenarios where nuances, such as negative statements, are relatively important for understanding the text (Aggarwal & Zhai, 2012).

e) **Stemming:** Stemming is an essential technique for NLP that reduces words to their roots or stems by removing suffixes and other endings. The goal is to standardize words that share the same semantic core, such as:

Original words: "atendido" (served), "atendimento" (service), "atender" (to serve)
Stemming to root form: "atend" (serve)
This standardization simplifies text analysis by reducing the dimensionality of the data and grouping similar terms together to improve the efficiency of analytical models (Manning, Raghavan & Schütze, 2008).

4.3 Data transformation

After preprocessing the data, we began the data transformation phase, which aims to adapt textual data into a more structured and suitable format for analysis. According to Manning, Raghavan, and Schütze (2008), machine learning and artificial intelligence models cannot directly process text in its original format, as they operate primarily with numbers and mathematical matrices. Therefore, it is necessary to transform texts into numerical representations so they can be used as input for classification and analysis algorithms. This process is known as vectorization.

Vectorization converts words, phrases, or documents into numerical formats that represent their features. Text-based machine learning models, such as sentiment classifiers, are better able to recognize mathematical patterns within vectors. As such, transforming textual data into vectors is essential to enable models to identify relationships, similarities, and semantic differences between words.

For the **ANE project**, we used the following approaches:

a) **Bag of Words (BoW):** Implemented using the CountVectorizer function from the scikit-learn library. This technique transforms texts into a word frequency matrix, where each row represents a sentence or document, and each column represents a unique word in the vocabulary. BoW is effective in identifying the presence or absence of specific words and works well in tasks that do not require complex semantic interpretation.

b) **Custom Tokenization:** This was performed using the TweetTokenizer from the nltk library, which was configured as a custom tokenizer in the CountVectorizer. This technique is especially useful for handling informal language found in comments or tweet-like texts, correctly processing hashtags, emojis, and abbreviations.

Simplified Example of Vectorization Used:

Corpus = ["Atendimento muito ruim" (Very bad service), "Atendimento excelente e rápido" (Excellent and fast service), "Péssimo atendimento" (Terrible service), "Atendimento muito bom" (Very good service)]

Expected Output – Generated Vocabulary:
{ 'atendimento': 0, 'muito': 4, 'ruim': 6, 'excelente': 2, 'rápido': 7, 'péssimo': 5, 'bom': 1}}

Each unique word in the corpus is assigned an index (its position in the vectorized form).

Copyright ©2025 by the authors. This conference paper is published under a CC-BY-4.0 license

For example:
atendimento → 0
muito → 4
ruim → 6

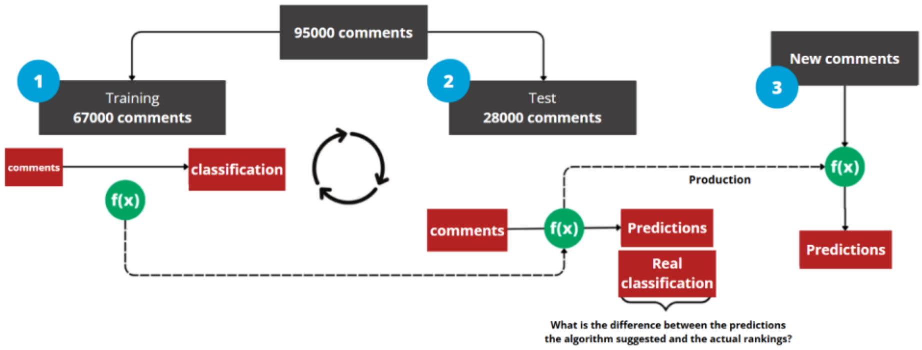
Vectorization of the sentence: "Atendimento muito ruim" (Service very bad)
[1, 0, 0, 0, 1, 0, 1, 0] #Atendimento muito ruim
The word atendimento (index 0) appears once.
The word muito (index 4) appears once.
The word ruim (index 6) appears once.
All other words have a frequency of 0 in this sentence.

The chosen approaches aim to address the characteristics of the analyzed comments, seeking to balance simplicity and effectiveness in pattern extraction. The combination of Bag of Words with custom tokenization proved to be efficient in handling the informal and unstructured nature of the texts.

4.4 Training and evaluation of classification algorithms

We used the traditional split to train and test the model, dividing the data into two subsets as shown in Fig. 2: training set (70%) – used to adjust the model parameters, and test set (30%) – reserved for evaluating the model's performance on unseen data.

Fig. 2 - Training and Test Split.



For the ANE project, we tested four algorithms to select the one with the best performance in the negative class. The tested algorithms were Logistic Regression, Support Vector Machines, Multinomial Naïve Bayes, and Random Forest. These algorithms are traditionally recognized for their efficiency in classification tasks, especially in multi-class categorization problems (McCallum & Nigam, 1998).

The models were evaluated using widely recognized classification metrics, such as precision, recall, F1-score, and accuracy. These metrics were applied to analyze the performance of the different algorithms in identifying patterns and correctly classifying comments within the dataset.

Table 2 presents the accuracy of the models tested for the comment classification task.

Tab. 2 - Overall accuracy of tested algorithms.

Algorithms	Accuracy
Logistic Regression	0.955577
Support Vector Machines	0.951323
MNB	0.950378
Random Forest	0.858223

Although the Multinomial Naïve Bayes (MNB) model did not achieve the highest overall accuracy, it stood out in a crucial aspect of our problem. The MNB model exhibited the highest precision and recall for the negative class, whereas the other models achieved approximately 90% accuracy in this class, which is the primary focus of this project. Accurately classifying negative comments—often characterized by informal language, abbreviations, or more subtle sentiments—was essential for choosing MNB.

Copyright ©2025 by the authors. This conference paper is published under a CC-BY-4.0 license

We initially experimented with the Gaussian Naïve Bayes (GaussianNB) variant, which assumes continuous feature distributions. However, since our input features were derived from vectorized textual data using TF-IDF (Term Frequency-Inverse Document Frequency)—a more suitable representation for discrete data—we also evaluated the Multinomial Naïve Bayes (MultinomialNB) algorithm. The MultinomialNB model performed better, particularly in handling the frequency distribution of words commonly found in natural language data. Consequently, we selected MultinomialNB as the Naïve Bayes proxy in the final comparison. The model was trained using its default hyperparameters: alpha = 1.0 for Laplace smoothing, fit_prior = True, and class_prior = None.

As observed in Fig. 3, the recall evaluation shows that the performance of the Multinomial Naïve Bayes model is slightly superior to the other tested algorithms.

Fig. 3 - General evaluation of the Multinomial Naïve Bayes model.

	What proportion of identifications were actually correct? PRECISÃO	How good is the model at predicting the class? RECALL	Balance between precision and recall F1-SCORE	Sample quantity SUPPORT
NEGATIVE	0.91	0.92	0.91	4385
NEUTRAL	0.84	0.12	0.20	643
POSITIVE	0.96	0.99	0.97	23937
accuracy			0.96	28965
macro avg	0.91	0.67	0.70	28965
weighted avg	0.95	0.96	0.95	28965

This result highlights the strength of MNB in handling text data with class imbalances or sparse distributions. Its probabilistic approach and sensitivity to word frequencies made it slightly more effective in identifying and recognizing patterns associated with negative sentiment. On the other hand, models like Logistic Regression and Support Vector Machines, despite achieving higher overall accuracy, were not as efficient in classifying this target class.

These findings reinforce the importance of aligning the model selection with the specific objectives of the analysis. For the ANE project, prioritizing the classification of negative feedback was essential, justifying the choice of Multinomial Naïve Bayes.

4.5 Algorithm Validation

The algorithm validation phase aimed to ensure that the developed model met the project objectives and the practical needs of monitoring Expresso services. This process was carried out in collaboration with the managers responsible for these operations to ensure that the chosen model provided acceptable metrics and produced coherent and useful classifications for decision-making.

For the managers' validation, an executable visual interface was developed that allowed analysis from the .csv file. The application enabled data to be locally loaded for classification and validation by the managers, based on the extracted file with evaluations from any unit and period. The local ANE - Análise EXPRESSO model allowed the comment classification in just a few seconds, highlighting those ones that required immediate attention.

The interface also included a feature for visualizing word clouds separated by class ('positive' and 'negative'), providing an overview of the most frequent terms and helping managers identifying patterns in the comments. In the negative comments (Fig. 4), the word cloud highlights recurring issues such as 'delay' or 'line', which may indicate critical areas that need to be prioritized. In the positive comments (Fig. 5), signs of efficiency or good service are quickly identified. This macro-level visualization of comments facilitates the identification of critical points to be analyzed.

Fig. 4 - Word Cloud with the Negative Class.

Copyright ©2025 by the authors. This conference paper is published under a CC-BY-4.0 license

Fig. 5 - Word Cloud with the Positive Class.

After the validation by managers and the approval of the model's efficiency in categorizing feedback, especially negative comments, fine-tuning adjustments were made to enhance alignment with the operational context. This included incorporating words and expressions commonly used by the team for service testing.

From that point, the script was adapted for integration with the Data Warehouse (DW) of the State of Goiás, enabling the classification of stored comments. To save and load the model, Python's **Pickle** module was used. However, since the model includes lambda functions, the additional **Dill** package was required (Izadi, 2019). Following these adaptations, the Pickle file was incorporated into a data processing routine, which now feeds a service monitoring dashboard. This dashboard, available exclusively to managers, utilizes the AI model developed to analyze citizen feedback, classifying their experiences with public services into predefined categories: positive, neutral, and negative. On average, 120 to 200 comments are processed and analyzed daily.

5. Discussion

This study highlighted the significance of applying NLP techniques to analyze comments related to the EXPRESSO services in the State of Goiás. The automated classification facilitated a deeper understanding of citizen feedback and enabled more efficient real-time monitoring, becoming an essential tool in overseeing the quality of public services.

A key point was the analysis of the neutral class. Although this class showed low accuracy due to the limited number of examples available in the training dataset, the model proved promising by effectively capturing neutral comments in new samples. This shows that, despite initial limitations, the model can recognize nuances in texts, indicating that the adopted approach is robust and has potential for refinement. A future improvement would involve expanding the dataset, especially by adding more examples of neutral comments, to reduce bias and improve performance in this class.

Another relevant aspect of the study was the geographical coverage of the database. The used data sample significantly represented the different regions of Goiás, which lends validity to the model in identifying the diverse contexts experienced by public service users. This representativeness is essential to ensure that the model's classifications take into consideration the needs of the entire population.

A graphical interface was developed to support the analysis of Excel and CSV files for local use by the management responsible for evaluations. However, it was also automated on the Goiás state server, being directly fed from the data warehouse, it is replicable. This opens up opportunities for the approach to be applied in other customer service models across Brazil, enabling a national expansion of the developed solution. The standardization of data

Copyright ©2025 by the authors. This conference paper is published under a CC-BY-4.0 license

input processes and the automation of classification make the model to be a scalable tool, which can be adapted to different regional contexts and the specific demands of each locality.

To successfully implement an artificial intelligence program focused on sentiment analysis of comments, it is essential to start the process with clear and transparent communication with managers, ensuring their support and alignment with the initiative's goals. It is also crucial to inform the team that AI is not here to replace them, but rather to eliminate repetitive and monotonous tasks, allowing them to focus on more relevant and strategic activities. The goal is to significantly enhance management by transforming large volumes of data into clear and actionable insights. With tools like sentiment analysis and word cloud visualizations, it becomes possible to quickly identify patterns, emotions, and recurring themes in customer or employee feedback. This enables managers to understand, objectively and in real time, how people perceive products, services, or even the organizational climate. With this information in hand, decisions are no longer based on assumptions but grounded in concrete data, increasing assertiveness, anticipating problems, and creating space for continuous improvement.

Finally, the results obtained reinforce the impact of using sentiment classification algorithms in public management. By prioritizing negative comments as the model's main focus, it allowed for more precise identification of critical points in the services provided, serving as an important monitoring tool for managers to understand citizens' improvements demands in public services. In this way, the work contributed to the improvement of sentiment analysis techniques in the public services context and to the enhancement of the quality of services offered to the population.

Acknowledgement

- **Funding or Grant:** this work is the result of a project developed by the Secretariat of Administration of the State of Goiás - SEAD. No specific monetary funding or grant number is associated with this project.
- **Data/Software Access Statement**: the data analyzed in this study are protected under Brazil's General Data Protection Law (*Lei Geral de Proteção de Dados - LGPD*, Law No. 13.709/2018) due to the presence of personal information from citizens. For further inquiries about the data, contact SEAD. The following software tools were used for data processing and analysis: Jupyter, Visual Studio Code, Python, and Microsoft Excel.
- **Contributor Statement**: Author 1: Conceptualization, Formal Analysis, Investigation, Methodology, Writing – Original Draft, Writing – Review & Editing; Author 2: Data Curation, Software, Validation, Visualization, Writing – Original Draft, Writing – Review & Editing; Author 3: Conceptualization, Project Administration, Supervision, Visualization, Writing – Original Draft, Implementation Manager, Writing – Review & Editing.
- **Use of AI**: During the preparation of this work, the authors used ChatGPT and Copilot to assist with translation between English and Portuguese. ChatGPT and Claude were also employed to support code generation and debugging. After using these tools, the authors reviewed, edited, and validated the content thoroughly, ensuring its quality and accuracy. The authors take full responsibility for the final content of this publication.
- **Conflict Of Interest (COI)**: There is no conflict of interest.

References

1. Aggarwal, C. C., & Zhai, C. (2012). Mining Text Data. Springer Science & Business Media.

2. Alpaydin, E. (2020). Introduction to Machine Learning. MIT Press.

3. Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press.

4. Liu, B. (2012). Sentiment Analysis and Opinion Mining. Morgan & Claypool Publishers.

5. Manning, C. D., Raghavan, P., & Schütze, H. (2008). Introduction to Information Retrieval. Cambridge University Press.

6. Pang, B., & Lee, L. (2008). Opinion Mining and Sentiment Analysis. Foundations and Trends in Information Retrieval, 2(1–2), 1–135.

7. Verma, S. (2022). Sentiment analysis of public services for smart society: Literature review and future research directions. Government Information Quarterly, 39(3), 101708. https://doi.org/10.1016/j.giq.2022.101708

8. Governo do Estado de Goiás. Lei nº 20.846/2020. Recuperado de

Copyright ©2025 by the authors. This conference paper is published under a CC-BY-4.0 license

https://legisla.casacivil.go.gov.br/pesquisa_legislacao/103353/lei-20846.

9. Royal Free London NHS Foundation Trust & DeepMind. "Our work on Streams." Recuperado de https://www.nhs.uk.

10. Internal Revenue Service. "IRS Artificial Intelligence in Tax Compliance." Recuperado de https://www.irs.gov.

11. E-Estonia. "X-Road: The Backbone of Digital Estonia." Recuperado de https://e-estonia.com.

12. National Health Authority of India. "Leveraging AI in Ayushman Bharat." Recuperado de https://www.pmjay.gov.in.

13. Chien, S., & Chen, W. (2020). AI in public administration: A framework for policy development and service innovation. Journal of Public Administration Research and Theory, 30(4), 567-583. https://doi.org/10.1093/jopart/muz048

14. Izadi, M. (2019, March 11). Pickle your model in Python. Medium. https://medium.com/@maziarizadi/pickle-your-model-in-python-2bbe7dba2bbb

15. SEAD - Secretaria de Estado da Administração. (2024, abril). Metodologia de avaliação dos serviços públicos: Gerência de avaliação da gestão e dos serviços públicos. SEAD.

16. Lara, R. D., & Gosling, M. S. (2016). Um modelo de gestão do relacionamento entre os cidadãos e a administração pública. Revista Eletrônica de Administração, 22(2), 352-376. https://doi.org/10.1590/1413-2311.0522015.59196

17. Goiás. (2020). Lei Estadual nº 20.846, de 2 de setembro de 2020. Institui a política de atendimento ao cidadão no Estado de Goiás e cria o Programa EXPRESSO. Diário Oficial do Estado de Goiás, Goiás.

18. Reichheld, F. F. (2003). The one number you need to grow. Harvard Business Review, 81(12), 46–54.

19. Russell, S., & Norvig, P. (2021). Artificial intelligence: A modern approach (4th ed.). Pearson.

20. O'Neill, R. (2021). AI in public administration: Current practices and future prospects. Public Administration Review, 81(3), 434-448. https://doi.org/10.1111/puar.13208

Copyright ©2025 by the authors. This conference paper is published under a CC-BY-4.0 license