# AUTOMATED TICKET CLASSIFICATION FOR TRAINING ISTAT'S PUC CHATBOT[1]

Samanta Pietropaoli, Gabriella Fazzi

**Abstract.** The increasing volume of user requests handled by ISTAT's contact center for supporting participants in official statistical survey has underscored the need for automated solutions to optimise ticket classification and reduce reliance on manual processing.
This study presents the development and evaluation of a supervised classification system that leverages Natural Language Processing (NLP) techniques to enhance the accuracy, efficiency and scalability of request management within a public administration context.
The proposed framework integrates a TF-IDF-based text representation with synthetic oversampling (SMOTE) and three supervised learning algorithms: Random Forest, LightGBM, and Multilayer Perceptron. The methodology also incorporates a tailored preprocessing pipeline—covering tokenisation, lemmatisation, stopword removal, and anonymisation of personal information—to ensure data quality and privacy compliance.
The classification system was designed to support the training phase of ISTAT's PUC chatbot, which will provide first-level assistance to citizens and establishments involved in statistical surveys. By generating high-quality labelled data, this approach aims to improve chatbot intent recognition and facilitate self-service interactions for survey respondents.
Model performance was evaluated using standard classification metrics, including accuracy and both weighted and macro-averaged F1 scores.
Among the tested configurations, LightGBM demonstrated the most balanced and robust performance. The results confirm the effectiveness of integrating machine learning and NLP into institutional workflows. Future work will explore the integration of the classifier into the generative architecture of Salesforce Agentforce, contributing to the evolution of intelligent support systems in citizen-facing public services.

## 1. Introduction

In recent years, the increasing demand for efficient and timely customer support has driven public institutions to explore innovative technological solutions. For the Italian National Institute of Statistics (ISTAT), this need is particularly pressing due to the vast volume of interactions with both citizens and businesses participating in official statistical surveys. Managing these requests manually is time-consuming and

[1] This article reflects only the opinions of the authors. This paper is the result of the joint work of the authors. S. Pietropaoli has written Sects 2, 3, 4, 5; G. Fazzi has written Sects. 1, 6;

prone to inconsistency, particularly when incoming messages vary in format, communication channels and language use.

To address these challenges, ISTAT is developing a first-level support chatbot within the framework of the "Punto Unico di Contatto" (PUC)[2], a unified omni-channel interface that allows external users to interact with the Institute in a standardised and integrated manner. The deployment of the chatbot is a key element within a broader digital strategy aimed at enhancing user experience, increasing operational efficiency, and reducing dependency on human operators for handling common and routine requests.

A chatbot is a virtual agent trained using specific machine learning algorithms to interpret and respond to user requests in a standardised and automated manner. Designed to simulate human-like conversations, it offers consistent assistance across a predefined set of tasks and queries.

The chatbot is trained on a set of *intents* that represent the various types of user queries—such as password recovery procedures, deadline requests, or survey clarification—which it must recognise in order to determine the appropriate response path. Accurate and comprehensive classification of past tickets is essential for generating high-quality labelled data that reflect real-world user needs. To address this, a robust classification pipeline that leverages advanced NLP techniques and supervised machine learning was developed. The goal is to automatically assign categories to the support tickets while simultaneously enhancing the quality of chatbot training data and improving the overall accuracy and efficiency of request management.

## 2. Context

The increasing relevance of Artificial Intelligence (AI) in the public sector is reflected in numerous research and innovation programs promoted at both European and national levels. For example, the AI4PublicPolicy project (Horizon 2020 - GA No. 101004480) developed a cloud-based platform to support policy development through AI technologies, including NLP and chatbots, with a strong emphasis on transparency and citizen feedback (European Commission, 2024). In particular, the Italian pilot led by the City of Genoa (Marceddu et al., 2024) closely aligns with this research's objective of automating the classification and analysis of helpdesk tickets using NLP models. Similarly, INPS—Italian National Social Security Institute— has placed AI at the heart of its digital transformation strategy (2022–2024). Key

[2] The PUC platform was launched in July 2024 and aims to centralise and harmonise interactions between ISTAT and its users by providing a single access point for information requests, survey participation support, and collaboration on statistical activities.

initiatives include AI-powered chatbots, automated contact center routing, classification of certified emails (PEC), and experiments with generative AI based on institutional knowledge bases (INPS, 2024). These efforts highlight the growing operational maturity of AI within Italian public administration. As a key actor in the implementation of the National Recovery and Resilience Plan (PNRR)[3]—especially under Mission 1, focused on digitalisation—INPS exemplifies how AI can enhance the quality and efficiency of citizen services.

The present research shares important parallels with these initiatives, particularly in its use of NLP for classifying user-generated content. It adopts a modular and experimental approach, focusing on fine-tuning pre-trained transformer models on domain-specific data, with special attention to data quality, linguistic challenges, and model explainability.

Such initiatives aim to drive digital transformation, increase efficiency in public service delivery, and promote citizen-centric governance through AI, data analytics, and cloud computing. While they mostly address large-scale service and policy transformation, this research contributes by adapting transformer-based NLP models to improve internal workflows at ISTAT.

These technological advancements also provide the broader context for the growing adoption of conversational agents in public administration. Among these, chatbots are increasingly adopted in public service delivery (European Commission, 2020), and are currently the most prominent form of AI applications in the public sector (Van Noordt & Misuraca, 2022). Typically adopted as a complementary service channel, chatbots are reshaping both information delivery and citizen interaction in public agencies (Vassilakopoulou et al., 2022). These systems are also recognised as tools for creating public value (Aoki, 2020), with implications for administrative processes, service quality, and the relationships between citizens and public institutions (Twizeyimana & Andersson, 2019). In this perspective, the ISTAT chatbot represents not only a technological advancement, but also a potential shift in how statistical services are accessed and perceived by users. Its long-term success depends not only on its initial design, but also on its ability to learn, adapt, and respond effectively to evolving user needs.

By automating the classification of helpdesk tickets, this research transforms cutting-edge AI and NLP methodologies into operational tools that support the PNRR's vision of a modern, efficient, and citizen-oriented public administration.

[3] The PNRR is Italy's program for using the European Union's Next Generation EU funds, with the aim of revitalizing the economy after the COVID-19 pandemic and promoting sustainable and inclusive growth.

## 3. Data and preprocessing

### *3.1. Dataset and classification taxonomy*

This research addresses the automation of the classification process by analysing over 24,000 historical support tickets, collected between June 2024—the launch date of the PUC platform—and February 2025, and categorised into 29 predefined classes. These tickets originate primarily from transcripts of phone calls handled by contact center operators and from email communications submitted by survey participants, specifically related to the activities of ISTAT's Data Collection Directorate. Most of these messages are written in Italian and show substantial variation in length, structure, and formality. To support the classification task, each ticket was pre-associated with a code[4] corresponding to one of 29 subcategories defined within a hierarchical, scenario-based taxonomy developed for the chatbot project.

*Scenarios* represent the overarching topical areas used to organise user inquiries and were established through a classification system specifically designed for this initiative. Initially informed by a 2020 experiment conducted in ISTAT on automated ticket classification[5], the structure was refined through in-depth analysis of frequently asked questions from various household and business surveys.

A multidisciplinary team reviewed and categorised these FAQs, leading to a hierarchical classification structure that distinguishes seven main scenario groups (labelled A to G), each grouping user requests by thematic domain and further divided into subcategories. These scenario groups are:

- A – Access and Navigation: Includes issues related to login credentials, website usability, and access to the data collection portals (A1–A10).
- B – Regulatory Aspects: Covers questions concerning mandatory response obligations, sanctions, privacy policies, and the legal framework for participation (B1–B8).
- C – Sample Composition and Variations: Encompasses requests about sample inclusion or exclusion, as well as participant status changes such as relocations, deaths, or business closures (C1–C8).

[4] The initial categorization of support tickets was generated during routine contact center operations. First-level operators handled cases by consulting a centralised Knowledge System—integrating FAQs and supporting documentation produced and maintained by Istat domain experts according to standardised procedures—and by accessing Istat's data acquisition back-office systems. As a result, ticket categorisation reflects shared operational practices and standardised information sources rather than an annotation scheme explicitly designed for machine learning purposes.
[5] For further information, see Bianchi et al., 2022.

- D – Interview Bookings: Refers to scheduling or rescheduling appointments with interviewers for CATI or CAPI surveys (D1).
- E – Assessment and Dispute Reports: Involves disputes or clarifications related to official notices, sanctions, or non-compliance assessments (E1).
- F – Questionnaire and Survey Information: Includes a broad set of queries about survey content, response deadlines, participation procedures, and technical aspects of questionnaire completion and submission (F1–F17).
- G – Communication with ISTAT: Covers inquiries about previous unanswered requests, official reminders, or clarification of institutional communications (G1–G4).
- H – Other: Reserved for tickets that do not fall into any of the predefined categories.

While the original taxonomy included 51 subcategories (as indicated in parentheses), it was streamlined to 29 for chatbot training purposes, maintaining the seven high-level scenarios. This restructuring preserved both the breadth of the thematic coverage and the clarity required for automated classification, while ensuring mutual exclusivity among classes.

### *3.2. Text Pre-processing*

To prepare the textual data for the supervised classification, a comprehensive and modular preprocessing pipeline was designed to ensure data quality, linguistic consistency, and full compliance with privacy regulations (e.g., GDPR). This phase was crucial to reduce data sparsity, eliminate noise, and prevent the model from inadvertently learning patterns based on sensitive information. The pre-processing steps, implemented using Python 3.11 and the SpaCy NLP library[6] (version 3.7), included the following components:

- Empty message removal: All records with missing values, null content, or messages composed solely of non-informative text (e.g., greetings or placeholders) were systematically excluded. This step was essential to minimize noise in the training data and avoid bias introduced by semantically empty inputs.
- Tokenization and lemmatization: Texts were segmented into individual tokens—the smallest units into which text is divided, which may range from words to subword components—using SpaCy's language model for Italian (`it_core_news_lg`). Each token was lemmatised to its canonical form to normalise lexical. These operations facilitate lexical generalisation and reduce

[6] SpaCy stands out for its support of over 70 languages through dedicated pipelines, many of which integrate advanced technologies such as BERT-based language models and multitask learning.

feature dimensionality, improving model robustness and generalisation, especially in low-resource or imbalanced class scenarios (Manning et al., 2010).

- Lowercasing: All tokens were converted to lowercase to treat words with identical but different casing as the same feature. This normalisation step prevents the artificial inflation of the vocabulary.
- Stopword removal: Standard Italian stopwords were removed using SpaCy's predefined list, which was supplemented with a custom lexicon of domain-specific terms and email-specific expressions. This included formal salutations, closings, and a curated list of recurring boilerplate phrases (e.g. legal disclaimers, environmental notices) and personal disclaimers, which were deemed semantically uninformative and irrelevant for the classification task.
- PII filtering: Personally identifiable information (PII)—including names, email addresses, telephone numbers, fiscal codes, and dates/time references—was detected and removed using a hybrid approach combining regular expressions and custom rule-based patterns in SpaCy. This process ensured full compliance with privacy standards and prevents the classifier from learning spurious associations based on sensitive identifiers. A similar rule-based strategy, successfully applied to Hungarian electronic health records, has demonstrated the effectiveness of NLP techniques for automated de-identification in domain-specific contexts (Berzi et al., 2023).

These pre-processing steps played a key role in reducing feature sparsity and noise, while standardising linguistic patterns across tickets and minimising the impact of irrelevant or sensitive information on model training. Such procedures are fundamental to improve computational efficiency and classification accuracy, especially when handling informal or semi-structured text typical of email-based communications (Méndez et al., 2005).

## 4. Methodology

The classification of support tickets was framed as a supervised multi-class problem, in which each message was assigned to one of 29 predefined categories. This approach follows established practices in text classification and relies on a pipeline integrating feature extraction, model training, and evaluation based on multiple metrics.

To convert the unstructured textual data into a machine-readable format, the input was vectorised using the Term Frequency–Inverse Document Frequency (TF-IDF) technique (Salton & Buckley, 1988), configured to extract both unigrams and bigrams. This allowed the model to account for individual word relevance as well as short lexical patterns observed in user communications. Tokens appearing in fewer than five documents were excluded to reduce noise and improve generalisation. A major challenge in this task was the strong class imbalance across ticket categories[7]. To mitigate this, all categories with fewer than 205 instances were merged into a single residual class (labelled X), thereby enhancing statistical robustness and improving model stability.

The dataset was then partitioned using an 80/20 stratified train–test split, with class labels numerically encoded. For oversampling, the Synthetic Minority Oversampling Technique (SMOTE) was applied to the training set to increase the representation of minority classes (Fernández et al., 2018). To limit oversampling noise, the number of nearest neighbours was dynamically adjusted based on the smallest class size (Chen et al., 2024). Three classification models were evaluated using the same pre-processing pipeline and TF-IDF feature set to ensure comparability. While the input representation and data split remained constant, the classifier architecture varied across model configurations. Evaluation metrics included overall accuracy, as well as class-level precision and recall, along with both weighted and macro-averaged[8] versions of precision, recall, and F1 scores[9].

The first setup employed a Random Forest (RF) classifier with 200 decision trees and a maximum depth of 50, balancing model complexity and overfitting risk. RF models aggregate predictions from multiple trees trained on random subsets of data, introducing stochasticity that reduces variance and enhances generalisation (Géron, 2019). This architecture is particularly suitable for capturing non-linear relationships and is widely recognised as a robust baseline for classification tasks (Breiman, 2001). The second setup retained the same pipeline architecture but replaced the classifier with light gradient boosting machine (LightGBM), a framework optimised for speed and memory efficiency in high-dimensional contexts such as text classification. The model was configured with 200 estimators and a maximum tree depth of 40, offering a trade-off between performance and generalisation. As in the RF model configuration, TF-IDF features—including both unigrams and bigrams—were used, and SMOTE was applied to address class imbalance. LightGBM's histogram-based learning algorithm enabled faster training and better handling of

[7] Some categories were significantly underrepresented in the original dataset, making it difficult for the classifier to learn consistent patterns. This included, for example, category A4 ("server malfunctions, slow system, supported browser, smartphone form filling"), F1 ("survey calendar"), and F11 ("form blocking, issues submitting the questionnaire").
[8] While weighted metrics account for class imbalance by giving more weight to frequent classes, macro-averaged metrics treat all classes equally, offering a better measure of the model's ability to generalise across minority categories.
[9] For a comprehensive description of these metrics, see Géron A., 2022.

sparse input, making it a strong candidate for large-scale, multi-class classification tasks (Ke et al., 2017). The third setup explored a neural network-based approach using a Multilayer Perceptron (MLP) classifier. The architecture consisted of a single hidden layer with 50 neurons and was trained for a maximum of 200 iterations, with early stopping enabled to prevent overfitting when no further improvement was observed on the validation set. As in the previous setups, the input was based on TF-IDF features. However, to improve convergence and reduce the dimensionality of the feature space, the vectorizer was slightly modified to extract only unigrams and to exclude tokens appearing in fewer than 10 documents. SMOTE was again applied to rebalance the training data. While the MLP offered a more flexible modelling structure, its performance was more sensitive to feature sparsity and class imbalance, particularly for underrepresented categories (Haykin, 1994).

## 5. Results

The outcomes of the three classification setups are presented and compared in Table 1. The evaluation focused on the models' ability to handle the imbalanced multi-class nature of the dataset, highlighting their respective performance across both frequent and underrepresented categories. Among the tested configurations, LightGBM yielded the best overall performance. It achieved the highest scores across all evaluation metrics, including accuracy (0.65), weighted precision, recall, and F1-score (all 0.65), along with the best macro-F1 score (0.52), indicating greater robustness in handling both frequent and rare classes. While weighted metrics account for class imbalance by giving more weight to frequent classes, macro-averaged metrics treat all classes equally, offering a better measure of the model's ability to generalise across minority categories. These results confirm LightGBM's suitability for high-dimensional, sparse classification tasks and its effectiveness in managing imbalanced data distributions.

The Random Forest classifier achieved slightly higher weighted precision (0.67) compared to LightGBM, suggesting a tendency to produce more confident positive predictions. However, this came at the cost of lower recall (0.61), which negatively impacted to overall balance of the model's performance (F1 = 0.63; macro-F1 = 0.50). This behaviour reflects a more conservative classification pattern, where relevant instances—particularly in minority classes—are more likely to be missed.

The MLP classifier, while conceptually more flexible due to its neural architecture, underperformed in comparison to the ensemble-based models. It yielded the lowest accuracy (0.57), as well as the lowest weighted and macro-averaged F1-scores (0.58 and 0.44, respectively). These results suggest that the MLP was more sensitive to the sparsity and imbalance of the input data. Despite preprocessing efforts to limit rare features and improve convergence, they struggled to generalise effectively, especially on underrepresented categories.

**Table 1 –** *Performance metrics of the classification models on the multi-class ticket dataset.*

| Model | Accuracy | Precision | | Recall | | F1 | |
|---|---|---|---|---|---|---|---|
| | | Macro | Weighted | Macro | Weighted | Macro | Weighted |
| Random Forest | 0.61 | 0.49 | 0.67 | 0.54 | 0.61 | 0.50 | 0.63 |
| LightGBM | 0.65 | 0.52 | 0.65 | 0.53 | 0.65 | 0.52 | 0.65 |
| MLPClassifier | 0.57 | 0.42 | 0.60 | 0.47 | 0.57 | 0.44 | 0.58 |

*Source: elaboration on support tickets from ISTAT's PUC platform (June 2024 - February 2025).*

**Figure 1 –** *Heatmap of class-wise F1-scores for Random Forest, LightGBM, and MLP classifiers on the dataset.*

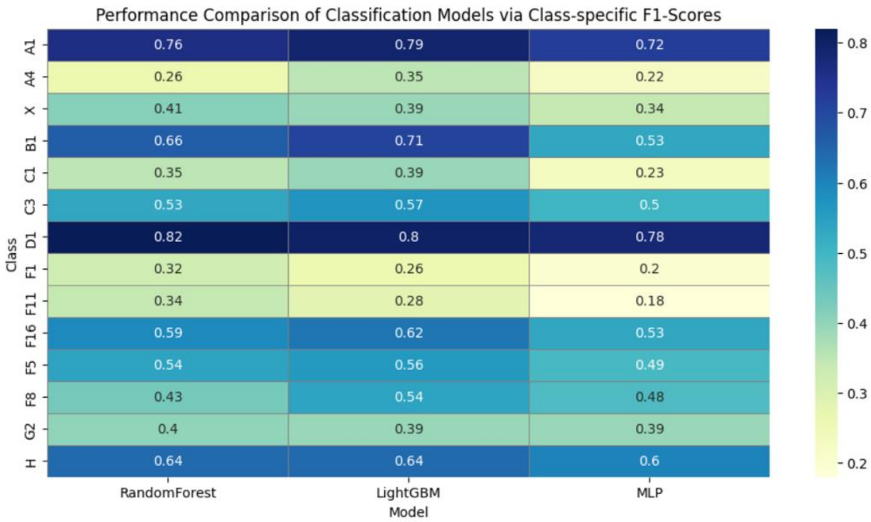


Figure 1 further illustrates these findings by reporting the F1-score achieved by each model for every target class. The most represented classes, such as A1 and D1, achieved high F1-scores across all models (above 0.7), indicating that the classifiers are able to correctly learn and predict these well-populated categories. Other classes such as F16, F5, and F8 showed moderate F1-scores (around 0.5–0.6), suggesting a fair level of predictive capability. In contrast, some underrepresented or imbalanced classes, including A4, F1, F11, and G2, exhibited significantly lower F1-scores (often below 0.4), highlighting the difficulty of all models in handling these categories. This result suggests that, despite the use of SMOTE with k_neighbors=5, oversampling may not be fully effective for classes with very few examples or complex feature distributions. Furthermore, LightGBM demonstrated improved performance over Random Forest for certain minority classes, such as B1 and C3, reflecting its stronger ability to capture class-specific patterns in the data. Conversely, the MLP classifier tended to underperform, particularly on less frequent classes—likely due to limited generalisation capacity or the need for further tuning or more complex architectures.

A qualitative inspection of misclassified instances revealed that most errors occurred between semantically adjacent categories within the same high-level scenario. Confusion was frequent between regulatory-related classes (e.g., B1 and B5) and questionnaire-related categories with overlapping procedural content (e.g., F5 and F8), indicating that misclassifications reflect genuine ambiguity in user language rather than randomness. Some failure cases involved very short or underspecified messages, such as brief requests lacking explicit references to the issue (e.g., "I need clarification about the survey"), difficult to disambiguate even for human operators. Other tickets contained multiple intents within a single message, leading the classifier to prioritise the most frequent or lexically salient category. These findings highlight intrinsic limitations of single-label classification in real-world support data and suggest that future work could explore multi-label approaches or the inclusion of contextual metadata to better capture user intent.

## 6. Conclusions and future perspectives

The experimental results confirmed the feasibility and effectiveness of using NLP and supervised learning techniques to automatically classify user support tickets in a public administration context. Among the tested models, LightGBM showed the most balanced and reliable performance, making it suitable for integration into real-world chatbot training pipelines.

Initially, development and testing were conducted using Einstein Bot, an AI solution embedded in Salesforce. This chatbot relied on NLP and ML technologies to analyse conversations, interpret user intent, and generate accurate responses through a supervised classification model.

During the project, ISTAT transitioned to a new conversational platform—Agentforce—representing a significant technological shift. Unlike traditional chatbots based on pre-defined decision trees, Agentforce is a generative agent capable of understanding natural language, reasoning through conversation, and dynamically selecting actions using both structured and unstructured sources.

Within this architecture, the classification model can enrich the Agent's knowledge base by assigning semantic labels to user requests, improving context interpretation, enhancing response relevance, and supporting real-time decision-making. However, some categories remain underrepresented or are associated with brief and ambiguous messages. These limitations can be mitigated by expanding the training dataset with additional contextual signals, historical behaviour, and newly collected tickets, which will enable finer-grained classification and a more detailed taxonomy.

Future work will focus on integrating the classification module into Agentforce and evaluating its impact on end-to-end chatbot performance. Ensuring that the system learns from high-quality, verified content is essential to reduce the risk of reproducing inaccurate responses.

This work demonstrates how AI-driven automation can enhance digital public services, providing a replicable model for intelligent support systems in data-intensive institutional environments. Within Agentforce, the classifier serves as an upstream semantic enrichment component: probabilistic intent labels guide response generation and knowledge retrieval, improving response consistency and reducing hallucination risks while posing challenges related to latency, confidence calibration, and taxonomy alignment.

## References

AOKI N. 2020. An experimental study of public trust in AI chatbots in the public sector. *Government information quarterly*, Vol .37, No.4.

BERZI A., BERÉNYI E., KÉPES Z., ANTAL B., VARGA Á. G., EMRI M. 2025. NLP-based removal of personally identifiable information from Hungarian electronic health records. *Frontiers in Artificial Intelligence*, Vol. 8, 1585260.

BIANCHI G., BELLINI G., BOSSO P., PAPA P. 2022. A machine learning based help-desk approach for units involved in official surveys. In *Proceedings of UNECE Expert Meeting on Statistical Data Collection 2022.*

BREIMAN, L. 2001. Random forests. *Machine learning*, Vol. 45, pp. 5-32.

CHEN,Y., ZOU J., LIU L., HU C. 2024. Improved Oversampling Algorithm for Imbalanced Data Based on K-Nearest Neighbor and Interpolation Process Optimization. *Symmetry*, Vol. 16, No. 3, p. 273.

EUROPEAN COMMISSION 2020. *AI watch - artificial intelligence in public services*. Technical report.

EUROPEAN COMMISSION 2024. *Horizon 2020 - Automated, Transparent Citizen-Centric Public Policy Making based on Trusted Artificial Intelligence*. Project web page. https://cordis.europa.eu/project/id/101004480/results.

FERNÁNDEZ A., GARCIA S., HERRERA F., CHAWLA N. V. 2018. SMOTE for learning from imbalanced data: progress and challenges, marking the 15-year anniversary. *Journal of artificial intelligence research*, Vol. 61, pp. 863-905.

GÉRON A. 2022. *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems*. O'Reilly Media, Inc.

HAYKIN S., 1994. *Neural networks: a comprehensive foundation*, Prentice Hall PTR.

INPS 2024. *Linee guida sull'implementazione di sistemi di Intelligenza Artificiale in INPS*. Direttiva del Direttore Generale n. 8 del 8 aprile 2024. Allegati circolare 14541.

KE G., MENG Q., FINLEY T., WANG T., CHEN W., MA YE Q., LIU T. Y. 2017. Lightgbm: A highly efficient gradient boosting decision tree. *Advances in neural information processing systems,* Vol. 30.

MANNING C. D., RAGHAVAN P., SCHÜTZE H. 2010. An introduction to information retrieval. In Mogotsi, I. C. (Ed.), *Information retrieval*, Cambridge University Press, pp. 192-195.

MARCEDDU A., Miccoli, M., Amicone, A., Marangoni, L., Risso, A. 2024. Artificial Intelligence for Urban Safety: A Case Study for reducing road accident in Genoa. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences,* Vol. 48, pp. 131-138.

MÉNDEZ J. R., IGLESIAS E. L., FDEZ-RIVEROLA F., DÍAZ F., CORCHADO J. M. 2005. Tokenising, stemming and stopword removal on anti-spam filtering domain. In: Marín, R., Onaindía, E., Bugarín, A., Santos, J. (Eds) *Current Topics in Artificial Intelligence*. CAEPIA 2005. Lecture Notes in Computer Science, Vol 4177. Springer, Berlin, Heidelberg.

SALTON G., BUCKLEY C. 1988. Term-weighting approaches in automatic text retrieval. *Information processing & management*, Vol.24, No. 5, pp. 513-523.

TWIZEYIMANA J.D., ANDERSSON A. 2019. The public value of E-Government–A literature review. *Government information quarterly,* Vol. 36, No. 2.

VAN NOORDT C., MISURACA G. 20222. Artificial intelligence for the public sector: results of landscaping the use of AI in government across the European Union. *Government information quarterly,* Vol. 39, No. 3.

VASSILAKOPOULOU P., HAUG A., SALVESEN L. M., PAPPAS I. O. 2023. Developing human/AI interactions for chat-based customer services: lessons learned from the Norwegian government. *European journal of information systems*, Vol. 32, No. 1, pp. 10-22.

Samanta PIETROPAOLI, Istat, pietropaoli@istat.it
Gabriella FAZZI, Istat, fazzi@istat.it