I.J. Modern Education and Computer Science, **2024, 5, 65-87**
Published Online on October 8, 2024 by MECS Press (http://www.mecs-press.org/)
DOI: 10.5815/ijmecs.2024.05.05

Information Technology for Gender Voice Recognition Based on Machine Learning Methods

Victoria Vysotska
Department of Information Systems and Networks, Lviv Polytechnic National University, Lviv, 79013, Ukraine
Osnabrück University, Osnabrück, 49076, Germany
E-mail: Victoria.A.Vysotska@lpnu.ua, victoria.vysotska@uni-osnabrueck.de
ORCID iD: https://orcid.org/0000-0001-6417-3689

Denys Shavaiev
Department of Information Systems and Networks, Lviv Polytechnic National University, Lviv, 79013, Ukraine
E-mail: denys.shavaiev.sa.2019@lpnu.ua
ORCID iD: https://orcid.org/0009-0004-0487-0780

Michal Greguš
Faculty of Managemen, Comenius University Bratislava, Bratislava, 82005, 25, Slovakia
E-mail: michal.gregus@fm.uniba.sk
ORCID iD: https://orcid.org/0000-0002-8156-8962

Yuriy Ushenko*
Department of Computer Science, Yuriy Fedkovych Chernivtsi National University, Chernivtsi, 58012, Ukraine
E-mail: y.ushenko@chnu.edu.ua
ORCID iD: https://orcid.org/0000-0003-1767-1882
*Corresponding Author

Zhengbing Hu
School of Computer Science, Hubei University of Technology, Wuhan, China
E-mail: drzbhu@gmail.com
ORCID iD: https://orcid.org/0000-0002-6140-3351

Dmytro Uhryn
Department of Computer Science, Yuriy Fedkovych Chernivtsi National University, Chernivtsi, 58012, Ukraine
E-mail: d.ugryn@chnu.edu.ua
ORCID iD: https://orcid.org/ 0000-0003-4858-4511

Received: 12 April, 2024; Revised: 10 June, 2024; Accepted: 24 July, 2024; Published: 08 October, 2024

Abstract: The growing use of social networks and the steady popularity of online communication make the task of detecting gender from posts necessary for a variety of applications, including modern education, political research, public opinion analysis, personalized advertising, cyber security and biometric systems, marketing research, etc. This study aims to develop information technology for gender voice recognition by sound based on supervised learning using machine learning algorithms. A model, methods and means of recognition and gender classification of voice speech samples are proposed based on their acoustic properties and machine learning. In our voice gender recognition project, we used a model built based on the neural network using the TensorFlow library and Keras. The speaker's voice was analysed for various acoustic features, such as frequency, spectral characteristics, amplitude, modulation, etc. The basic model we created is a typical neural network for text classification. It consists of the input layer, hidden layers, and the output layer. For text processing, we use a pre-trained word vector space such as Word2Vec or GloVe. We also used such techniques as dropout to prevent model overtraining, such activation functions as ReLU (Rectified Linear Unit) for non-linearity, and a softmax function in the last layer to obtain class probabilities. To train a model, we used the Adam optimizer, which is a popular gradient descent optimization method, and the "sparse categorical cross-entropy" loss function, since we are dealing with multi-class classification. After training the model, we saved it to a file for further

use and evaluation of new data. The application of neural networks in our project allowed us to build a powerful model that can recognize a speaker's gender by voice with high accuracy. The intelligent system was trained using machine learning methods with each of the methods being analysed for accuracy: K-Nearest Neighbours (98.10%), Decision Tree (96,69%), Logistic Regression (98.11%), Random Forest (96.65%), Support Vector Machine (98.26%), neural networks (98.11%). Additional techniques such as regularization and optimization can be used to improve model performance and prevent overtraining.

Index Terms: Voice recognition, Machine learning, Gender recognition, Information technology, Cybersecurity, Authentication, Natural language processing, neural networks, Gender classification models, Voice-by-sound recognition

1. Introduction

Voice-by-sound recognition is an urgent problem in the field of modern education, cybersecurity, artificial intelligence, computational linguistics and natural language processing (NLP), and computer science, including in modern education and distance learning, for example, during the COVID pandemic or wars in the presence of poor online communication. Recent years have seen significant development in voice technologies and applications such as voice assistants, automated call systems, speech recognition technologies, etc. These information technologies (IT) are becoming increasingly popular in various industries, including the development of consumer devices, the automotive industry, modern education and distance learning, medicine, cybersecurity, telecommunications, communications, and many others. Voice recognition, including gender recognition of the speaker, is an important component of this IT, as it allows computers and systems to understand and interpret voice commands and user messages. This has a significant impact on the convenience and efficiency of human-computer interaction. Voice-by-sound recognition also has great potential to open new opportunities, such as developing applications for people with special needs (for example, in modern education and distance learning), improving security systems, developing innovative interfaces, and much more. Therefore, research and development of effective gender recognition algorithms based on voice-by-sound analysis is an important task in modern computer linguistics, artificial intelligence, cybersecurity and computer science.

A large number of scientists around the world are currently dealing with the problem of gender identification by analysing posts on Internet networks, in particular, many studies are focused in the direction of gender linguistics. Given that language is an important means of self-expression, several characteristics of text may indicate possible gender differences in the way we communicate. Thanks to machine learning and natural language processing technology development, the possibilities of gender detection are increasing. This data can be used to analyse and understand the behaviour of users on the Internet, identify and forecast trends in the consumer market, as well as to create more effective strategies for communication and interaction with the audience. The use of such methods allows to avoid stereotypes and hidden associations, as well as to ensure the representativeness of the analysis according to gender criteria. The growing attention to issues of equality and diversity also makes this area relevant, as it can be used to analyse and identify possible stereotypes, discrimination or inequality that may exist in online communities. Accordingly, the detection of gender with the help of artificial intelligence has great potential in many areas, which makes the direction of research relevant in today's information society. A large volume of structured data when solving a problem requires a mandatory stage of data modelling of the database, and application software implementation for effective work requires the use of an object-oriented approach. Automation of gender identity detection based on support vector machine (SVM) analysis of Internet posts will contribute to the creation of safe and tolerant web environments.

Usually and most often, the detection of gender from posts on social Internet networks using NLP is performed to analyse and classify text data to determine gender, which is carried out on the features basis of language and writing style that are characteristic of certain groups of users. The corresponding method works by transforming the input data in the form of a set of trained gender classifiers and vectorizers used during training and the post of social Internet networks for analysis into the output data in the form of an assessment of whether the post for analysis belongs to a certain gender. But not all women have, so to speak, a feminine type of speech, and, accordingly, men have a masculine type of speech. Therefore, interesting, relevant and promising research will be the analysis of speech through the properties and acoustic characteristics of the sound of the voice, timbre, frequency, pitch, etc. The input data of the method is a set of gender classifiers trained on English-language data (because there are many freely available datasets specifically for this language) and the corresponding vectorizers used during training and the post of social Internet networks for analysis. The zero step is to select the gender type to analyse and load the appropriate model of the gender classifier with the vectorizer. The first step is to pre-process the social mass media post, which includes removing stop words and stop characters, as well as checking the writing language. If the post language is not English, the accuracy of recognition is unfortunately not guaranteed. The next step is the vectorization of the pre-processed post of social Internet networks, after which the step of classification by a trained gender classifier takes place. The fourth step is the formation of conclusions for the user of the post on social Internet networks. The initial data of the method is an

assessment of the post's belonging to the specified gender for analysis. Another problem is the collection of data (formation of datasets) of voice recordings for conducting similar research on voice recognition and especially its use in biometric systems. This is due to ethical considerations and the security of users' conference data (for example, to create voice deep fakes or voice phishing from alleged acquaintances or relatives, etc.), related to privacy, consent and bias. This is especially important when dealing with sensitive data such as voice recordings.

The goal of the study is to develop an effective voice-by-sound recognition model capable of accurately classifying male and female voices. The project includes several specific tasks:

1) Uploading and preparing a dataset containing acoustic features of male and female voices.
2) Extracting features from voice samples, such as frequency, intensity, formants, etc.
3) Splitting the dataset into testing and training datasets.
4) Development and training of a machine learning model (Support Vector Machine, Logistic Regression, K-Nearest Neighbours, Decision Tree, Random Forest, neural networks) for voice classification.
5) Assessing model performance using such metrics as F1-score, recall, precision, and accuracy.
6) Visualization and discussion of results, for example, the creation of graphs and diagrams for comparing the distribution of acoustic features in male and female voices.

Project tasks reflect the sequence of steps necessary to achieve the goal of developing a voice recognition system:

1. Collecting voice recordings: To record male and female voice samples for further processing and analysis. This may require the use of special sound recording devices and software.

2. Processing of voice recordings using the R programming language: To analyse voice recordings using the R programming language. This may involve identifying those acoustic features that are representative of male and female voices. Possible tasks may include the analysis of the frequency spectrum, sound intensity, time characteristics, etc.

3. Creating numerical voice scores: Based on the results of processing voice recordings using R, develop an algorithm for generating numerical scores that will reflect the differences between male and female voices. These scores may be based on statistical metrics, machine learning models, or other methods of analysis.

4. Using Python for voice classification: Using Python, to develop machine learning models that will be trained on numerical voice scores created at the previous step. These models will be used to classify voices as male or female.

5. Training the model to classify voices: Using the set of recorded voices, which includes data on male and female voices, to train the machine learning model using the training dataset. The training will help the model identify male and female voice patterns and specificities.

6. Validation and assessment of the model: To check the performance of the trained model using the test dataset to evaluate its accuracy and reliability in terms of gender voice recognition. To use assessment metrics, such as accuracy, precision, recall, etc. to assess model performance.

7. Using the developed model: The practical value of this project consists in the possibility to use the developed model for distinguishing between male and female voices. It can be useful in multiple fields, for example, automatic voice command recognition, voice identification for authentication systems, voice data analysis for linguistic research, and others.

The object of the research is the process of voice analysis and gender recognition based on acoustic characteristics. The result of the research is the development of a model or an algorithm that is capable of differentiating between a male and a female voice with high accuracy. The resulting model can be used for voice recognition and other applications where it is important to identify the speaker’s gender based on voice. The research subject is models, methods, and means of sound processing as well as gender recognition by parametric voice analysis based on machine learning. Voice analysis reflects the process of extracting and processing the acoustic parameters that characterize the voice signal.

The scientific novelty of the study consists in the development, based on machine learning, of acoustic models and methods, which allow obtaining key characteristics of the voice and using them for gender recognition. The use of these models and algorithms makes it possible to achieve high accuracy in determining the speaker’s gender. It also allows automating the process of determining gender and makes it faster and more efficient. The specific feature of this project is using the R programming language to process voice recordings and create numerical voice scores and the Python programming language to classify the voice as male or female. This combination of different tools and methods provides a new approach to voice analysis and gender recognition. As a result of this project, new approaches to voice analysis and gender recognition will be obtained, which may have important scientific and practical applications for voice recognition. The developed project has practical value in certain fields of application. The following are some of the areas where this project can be useful:

1) Medicine: Human voice can reflect certain characteristics of health, for example, changes in the voice can be an indicator of certain diseases or pathologies. The developed project can be used to diagnose and monitor certain medical conditions through voice analysis.

2) ***Biometric systems***: The results of this project can be used for voice recognition and person authentication. This can be used, for example, in access control systems or person identification systems.
3) ***Commerce***: Voice analysis can have practical applications in commercial projects such as marketing research or advertising campaigns. Voice gender recognition can help improve personalized advertising strategies or understand consumer preferences.
4) ***Voice research***: This project can be used to study the human voice and its features in general. It can help understand correlations between different acoustic characteristics of the voice and gender, which can lead to discoveries in the field of linguistics and phonetics [61-63].

2. Literature Review

Determining the age and gender of a person can be used as auxiliary information when solving several problems [[12]]. Advertising and dialogue help systems used in telephones are more effective if they take into account people's age and gender [[1], [23]]. The knowledge of a person's gender and age can be used as additional information for identity authentication and verification as well as for word recognition in speech. However, the known methods of determining the age and gender of people are not sufficiently reliable under the conditions of various interferences [[37]]. It can also be an additional parameter in speaker recognition [[3], [36], [60]]. Speaker recognition is the identification of a person based on acoustics voice characteristics. It is used as an additional parameter in speaker identification (for instance, during identification/authentication as an extra component of verification), but rarely during speech recognition (except to whom an unknown fragment of speech can potentially belong, for example, when intercepting telephone conversations, etc.) [[3], [36], [37], [60]]. Acoustic models reflect behavioural biometrics based on the speaker's anatomy (for instance, the size and shape of the throat or the mouth) and the specifics of acoustic parameters (for instance, the main tone of the voice, and the speaking style).

With the development of technology, many companies switched from traditional to modern biometric authentication methods [[30]]. Palm, facial and fingerprint recognition remain the most common types, but voice recognition is also growing in popularity [[8]]. Biometric voice recognition uses unique characteristics of the human voice to identify a person, which provides additional protection of personal data and finances and allows a user to abandon passwords that require physical input [[9], [15], [27], [58]-[63]]. Biometric voice recognition has several options for use [[10], [14], [43], [44]].

- ***Contact centre*.** This is the most common example of using voice biometrics. When interacting with customer service, using voice for authentication saves effort and time for both the service provider and the consumer.
- ***Fraud detection*.** Voice biometrics is a powerful fraud detection application because it offers a secure and tamper-proof authentication mechanism. Thus, this tool is used to prevent unauthorized access to the customer's data or finances.
- ***Financial services*.** The world market of financial services has experienced significant changes. Customers' lives have become easier thanks to online banking and different financial technologies. Financial institutions use voice biometrics for more convenient, faster, and safer interactions with customers.
- ***Digital signatures*.** Voice biometrics can be used to establish voice signatures used to sign documents such as life insurance contracts. Transactions can also be authorized using voice signatures.
- ***HR*.** The voice biometrics use in HR applications is already commonplace. This technology is a secure alternative to badging systems for organizations with large staff.

Advantages of voice biometrics [[8], [9], [15], [27], [30], [58]]:

1. ***Low operating costs.*** Even banks and call centres can save money by using voice authentication. This saves millions of euros by eliminating many of the steps required in standard authorization methods.

2. ***Improved user interaction***. Another benefit of voice biometric system is that they can significantly improve interaction with users. Subscribers no longer need to enter passwords, and PIN codes or answer verification questions to verify their identity.

3. ***Accuracy and reliability***. Voice authentication is more reliable and accurate than passwords, which can be easily forgotten, changed or guessed by fraudsters. It is unique like fingerprints.

4. ***The technology is easy to implement***. Many companies value voice recognition technology in terms of usage and adoption. Some biometric technologies can be difficult to integrate into a business, but voice biometric systems can usually be implemented without additional equipment and require few resources.

Also, the problem is the collection of data (formation of datasets) of voice recordings for conducting similar research on voice recognition and especially its use in biometric systems. This is due to ethical considerations and the security of users' conference data (for example, to create voice deep fakes or voice phishing from alleged acquaintances or relatives, etc.), related to privacy, consent and bias. This is especially important when dealing with sensitive data such as voice recordings.

Finally, artificial intelligence can be used to create gender-inclusive technologies [[5], [24]]. AI can be used to develop technologies that are designed to be gender-neutral and not support gender stereotypes [[16], [35], [39]]. For example, AI can be used to develop voice recognition technologies that are designed to recognize both male and female voices [[7], [50], [51]].

AI has the potential to become a powerful application in the fight for gender equality [52-54]. By using artificial intelligence to analyze data, identify gender biases and create gender-responsive technologies, we can make progress towards gender equality [55-64]. Also, voice biometrics can be used for male or female voice imitation in speech synthesis systems for people who have lost the ability to speak [[22]]. It can be used for the artificial dubbing of film characters and the dubbing of videos without human dubbing. For speech synthesis, you can mix male or female voices with different timbres or accents. This is the list of 2 analogues to our product under development: SpeakerRecognitionAPI; Genderize.io; Pindrop; Nuance VocalPassword; and VoiceIt. The main common element of these analogues is their ability to recognize and classify voice gender. Each of the products has its features and its approach to gender recognition by voice. The analogues differ in their functionality, accuracy, availability, and used algorithms:

1. ***Speaker Recognition API***. This API provides the ability to recognize and identify human voices, including the speaker's gender. Advantages: easy integration, high accuracy, scalability. Disadvantages: dependent on audio quality: voice recognition results can be sensitive to the quality of audio input. Poor recording quality can lead to less accurate results and limit the algorithms, which can affect the accuracy and overall performance of API. Basic methods of voice processing: acoustic voice analysis, machine learning, statistical models, and deep learning, which allows the detection of complex dependencies in the voice automatically and provides high accuracy of voice recognition by gender.

2. ***Genderize.io*** is an API service, designed to identify the gender of the speaker by name, but it also provides the possibility to define gender by voice. It uses a deep learning model to analyse voice data and identify the speaker's gender. Main advantages: easy to use, breadth of coverage, voice recognition support. Disadvantages: dependent on names and limited accuracy. Despite extensive database coverage, the accuracy of Genderize.io can vary, especially in the case of rare or little-known names. Basic methods of voice processing: acoustic voice analysis, use of statistical models, and machine learning algorithms. Algorithms can learn from a large set of voice recordings with a known gender and use that knowledge to classify new voice data.

3. ***Pindrop*** is a company that specializes in voice recognition and voice authentication based on voice biometrics and voice analysis technology for personal identification and voice recognition. Advantages: high accuracy of voice recognition, fraud protection, real-time support. Disadvantages: limitations on the quality of voice recording, and the background noise influence. Voice processing methods: deep neural networks and machine learning algorithms. Pindrop algorithms analyze various voice parameters, including timbre, frequency response, intonation, and rhythm, to create a voice profile of a person.

4. ***Nuance VocalPassword*** is a system developed by Nuance Communications, which specializes in voice biometrics and voice authentication. Uses unique voice characteristics to identify and authenticate a person. Advantages: high accuracy of voice recognition, easy to use, high security. Disadvantages: dependent on recording quality and requires training, which may take some time and resources. Uses sophisticated voice processing algorithms, including spectral voice analysis, formant identification, intonation and rhythm detection, and other acoustic parameters to create a unique voice profile of a person.

5. ***VoiceIt*** is a voice recognition system designed to authenticate users by voice. It uses voice biometrics for identity and user authentication and provides real-time voice recognition via an API interface. VoiceIt supports different languages and accents, which makes it usable in different countries and environments. Advantages: high accuracy of voice recognition, integration flexibility, and open access to documentation. Disadvantages: limits depending on subscription plan and the quality of recordings. VoiceIt uses voice processing algorithms that include spectral analysis, formant detection, speech feature extraction, and other acoustic parameters. The system can also use machine learning and neural networks to improve the voice recognition process.

Gender and age classification of people can be used as features of a set of heterogeneous evaluations of the speech signal based on models of hearing and speech production. The vector of features for gender and age classification includes Mel-frequency cepstral coefficients (MFCCs) which can be rationally combined with momentary functions of the fundamental tone frequency and formant features. Mel is a unit of the frequency measurement of a certain sound, created after statistical processing of a large number of data on the perception of the pitch of signals. The Mel scale is an empirical scale based on human perception of sound frequency. Based on MFCC, colour features are calculated for neural networks when recognizing a specific voice command. The input digital signal is divided into time segments of the same duration, which allows us to obtain a set of estimates of features that change over time. Segmentation is then performed, which includes making a decision to detect the speech signal and finding the time limits of the beginning and end of the speech. This makes it possible to exclude fragments that do not carry linguistic information at the stage of recognition. The fundamental tone frequency distribution features can be used to recognize the gender and age of announcers. Estimates of the pitch frequency for male announcers had an average frequency of 128 Hz with a range of possible values from 58 Hz to 238 Hz, and for women, the average frequency was 256 Hz with a range of 135 Hz to 522 Hz. The histograms found from the set of time-consecutive estimates of the frequency of the fundamental tone are

asymmetrical - to their fashion: in women's voices, the slope is steeper from the side of small periods than for long periods, while the reverse pattern is observed for men's voices. To recognize gender and age, it is rational to use normalized central moments of the frequency of the fundamental tone up to the 6th order inclusive. Different classifiers can be used to classify people by gender and age. These can be Gaussian mixture models, k-nearest neighbours (kNN), SVM, and support vector regression (SVR). Solving this problem requires the formation of a database of class standards. Application as MFCC features, analysis of acoustic features of sound and formant frequency estimates combined with estimates of the frequency of the fundamental tone and its normalized moments made it possible to achieve a high percentage of accuracy for correct gender recognition. Basic and common acoustic features of sound for statistical analysis and machine/deep learning are [modindx, meanfreq, dfrange, sd, maxdom, median, mindom, q25, meandom, q75, maxfun, iqr, minfun, skew, meanfun, kurt, peakf, sp_ent, centroid, sfm, mode [1-2].

3. Material and Methods

The main task of this project is to develop a system for voice gender recognition. The goal of the project is to create a tool capable of identifying a person's gender based on the features of their voice. By comparing our application with its analogues, it is possible to determine how competitive it is when it comes to gender recognition by voice. The advantages of our application are the following:

1. ***High accuracy in gender recognition***. By using powerful algorithms and neural networks, our app can achieve high accuracy when identifying gender by voice.
2. ***A wide range of supported languages and accents***. Our app can recognize voices in different languages and accents, making it universal and suitable for use in different countries and environments.
3. ***Fast processing and recognition***. Our application offers effective voice processing algorithms for fast recognition of the speaker's gender by voice.
4. ***User-friendly interface and easy to use***. Our application can be easily integrated into various platforms and systems.

Disadvantages of our competitors:

1. Low accuracy of speaker gender recognition by voice, can lead to incorrect results.
2. Support of fewer languages and accents, which limits their applicability across cultures and geographic regions.
3. High cost, which makes them unavailable to some users or organizations

Given the advantages of our application, including its high accuracy, wide support for languages and accents, fast processing, and user-friendly interface, we can compete with peers and attract the attention of users and enterprises who are looking for a reliable and effective voice gender recognition tool [[11], [57]].

To compete with analogues, it is necessary to solve some problems:

1. Accuracy of recognition: The application should achieve high accuracy in identifying gender by voice, avoid false results, and ensure reliable results regardless of the different features of speakers' voices.
2. Universality: The application should be able to recognize voice gender in different languages and accents, and work with different types of voice devices. It must be adapted to different cultural and linguistic contexts to have a wide range of applications and meet the needs of different users.
3. Speed and performance: The application should provide fast processing of voice data and instant output of results. It should be optimized to work efficiently even with a large number of voice recordings.
4. Privacy protection: The application should comply with privacy requirements. It should ensure the security and privacy of voice data, and prevent its misuse or disclosure.

Each musical sound is characterized by the following acoustic properties: pitch, strength, timbre, duration, etc. In spoken language, the frequency of the main tone of the voice in men varies from 85 to 200 Hz, and in women from 160 to 340 Hz. Modulation of the voice by pitch ensures the expressiveness of oral speech. The concept of tonal range is distinguished, that is, the ability to produce sounds within certain limits, from the lowest tone to the highest. These possibilities are individual for each person. The singing voice has a large range.

In spoken language, the intensity of the voice is from 40 to 70 dB, the voice of singers has 90 to 110 dB, and in some cases, it can reach 120 dB (noise power of an aircraft engine). Human hearing has adaptive capabilities, thanks to which you can listen to quiet sounds against a background of loud ones, or gradually get used to noise and begin to distinguish sounds. However, even with this, loud sounds are not indifferent to human hearing - at 130 dB, the pain threshold occurs, 150 dB is intolerable, and 180 dB is fatal for a person.

To characterize a normal voice, there is such a concept as a tonal range - the volume of the voice - the ability to produce sounds within certain limits from the lowest tone to the highest. This is a property for each person individually.

The tonal range of a speaking voice in women is within one octave, in men it is slightly less, that is, the change in the main tone during a conversation, depending on its emotional colour, varies within 100 Hz. The tonal range of the singing voice is much wider - the singer must have a voice in two octaves. Famous singers whose range reaches four and five octaves: can take sounds from 43 Hz - the lowest voices - to 2,300 Hz - high voices.

A person's voice can vary in pitch within about two octaves. 4-6 tones are enough for normal conversational speech. Voice range can vary greatly from person to person. The frequency characteristics of the main voice types are as follows: for men 80-340 Hz (Bass), 96-426 Hz (Baritone) and 128-512 Hz (Tenor), and for women 170-680 Hz (Contralto), 216-864 Hz (Mezzo-soprano) and 256-1024 Hz (Soprano). The voice range in children is much smaller than in adults and increases with age almost equally in girls and boys: 320-512 Hz for 8-10 years, 290-580 Hz for 10-12 years and 256-680 Hz for 12-14 years Both girls and boys have higher voices (treble) and lower voices (alto).

To achieve this goal, we plan to use the sound data set that will include voice recordings of women and men [[1]-[2], [18], [20], [34], [56]]. From these data, we will extract various acoustic characteristics of a specific speaker. C_s from the corresponding fragment of the sound file with the voice of this speaker, such as frequency characteristics, spectral characteristics, amplitude, modulation, etc. [[13], [17], [19], [45], [52], [55]]:

$$C_s =< c_{mode}, c_{median}, c_{sd}, c_{meanfreq}, c_{centroid}, c_{IQR}, c_{Q25}, c_{peakf}, c_{Q75}, c_{sfm}, c_{sp_{ent}}, c_{kurt}, c_{skew}, c_{modindx}, c_{dfrange}, c_{minfun}, c_{mindom}, c_{maxfun}, c_{maxdom}, c_{meanfun}, c_{meandom} >,$$

where c_{mode} is mode frequency; c_{median} is median frequency (kHz); c_{sd} is standard frequency deviation (median frequency in kHz); $c_{meanfreq}$ is mean frequency (kHz); c_{IQR} is interquartile range (kHz); $c_{centroid}$ is frequency centroid; c_{Q25}, and c_{Q75} are first and third quartile (kHz); c_{peakf} is pea frequency (with the highest energy); c_{sfm} is spectral uniformity; c_{sp_ent} is spectral entropy; c_{kurt} is kurtosis; c_{skew} is asymmetry or inequality; $c_{modindx}$ is modulation index, which is calculated as the accumulated absolute difference between adjacent fundamental frequency measurements divided by the frequency range; $c_{dfrange}$ is the dominant frequency range measured indirectly in the acoustic signal; as well as c_{minfun} and c_{mindom} is the minimum value of the fundamental/dominant frequency, respectively, measured indirectly in the acoustic signal; c_{maxfun} and c_{maxdom} are the maximum value of the fundamental/dominant frequency, respectively, measured indirectly in the acoustic signal; $c_{meanfun}$ and $c_{meandom}$ is the average value of the fundamental/dominant frequency, respectively, measured indirectly in the acoustic signal.

The module for the speaker's acoustic features analysis is presented as a tuple:

$$M_i^a =< C_s, \alpha, \beta, \gamma, A_i, P >,$$

where $\alpha(x, y)$ is logical relation operator for comparison $x < y$; $\beta(x, y)$ is logical relation operator for comparison $x \geq y$; $\gamma(\alpha, \beta, C_s)$ is output formation operator (the output is a conclusion on the speaker's gender) based on production rules P; A_i is the result of a logical conclusion (male or female) on the *i*-th fragment of a recording of the speaker's voice.

Based on the data from [[1]] (1,584 female voices and 1,584 male voices), the average, maximum and minimum values of each of the acoustic characteristics of the voice recordings were selected. Table 1-2 shows examples of acoustic characteristics for random three male voices and three female voices. As can be seen from the data of only 6 people, some indicators differ significantly in range depending on the speaker's article. For example, c_{mode} for these random male voices are 0, and for female voices, the average value is 0.162636517. The value of c_{maxdom} for men ranges from [0.007813; 0.054688], and for women - [0.554688; 3.59375]. The value of c_{minfun} for men ranges from [0.015656; 0.015826], and for women - [0.034483; 0.062257]. The range of c_{maxfun} values for both sexes coincided. Based on the results of the analysis of 3,168 recorded voice samples, it was determined, for example, that at c_{mode}<0.14 or c_{mode}>=0.059 the voice will be male, and at c_{mode}<0.012 it will be exactly female. If the voice is at c_{mode}<0.14 and c_{mode}>=0.012, then it is better to analyze c_{maxdom} as well. If c_{maxdom}<6.6, then the voice is male.

Table 1. Examples of acoustic characteristics for three random male voices

#	Acoustic characteristics	Male 1	Male 2	Male 3	Average value	Min value	Max value
1	meanfreq	0.0597809849598081	0.066008740387572	0.0773155026958227	0.067701743	0.059781	0.077316
2	sd	0.0642412677031359	0.0673100287952527	0.0838294209445061	0.071793572	0.064241	0.083829
3	median	0.032026913372582	0.040228734810579	0.0367184586699814	0.036324702	0.032027	0.040229
4	Q25	0.0150714886459209	0.0194138670478914	0.00870105655686762	0.014395471	0.008701	0.019414
5	Q75	0.0901934398654331	0.0926661901358113	0.131908017402113	0.104922549	0.090193	0.131908
6	IQR	0.0751219512195122	0.0732523230879199	0.123206960845246	0.090527078	0.073252	0.123207
7	skew	12.8634618371626	22.4232853628204	30.7571545800584	22.01463393	12.86346	30.75715
8	kurt	274.402905502067	634.613854542068	1024.927704721	644.6481549	274.4029	1024.928
9	sp.ent	0.893369416700807	0.892193242265734	0.846389091878782	0.87731725	0.846389	0.893369
10	sfm	0.491917766397811	0.513723842537073	0.478904979116727	0.494848863	0.478905	0.513724
11	mode	0	0	0	0	0	0

12	centroid	0.0597809849598081	0.066008740387572	0.0773155026958227	0.067701743	0.059781	0.077316
13	meanfun	0.084279106440321	0.107936553670454	0.0987062615673936	0.096973974	0.084279	0.107937
14	minfun	0.0157016683022571	0.0158259149357072	0.0156555772994129	0.01572772	0.015656	0.015826
15	maxfun	0.275862068965517	0.25	0.271186440677966	0.265682837	0.25	0.275862
16	meandom	0.0078125	0.00901442307692308	0.00799005681818182	0.008272327	0.007813	0.009014
17	mindom	0.0078125	0.0078125	0.0078125	0.0078125	0.007813	0.007813
18	maxdom	0.0078125	0.0546875	0.015625	0.026041667	0.007813	0.054688
19	dfrange	0	0.046875	0.0078125	0.018229167	0	0.046875
20	modindx	0	0.0526315789473684	0.0465116279069767	0.033047736	0	0.052632

Table 2. Examples of acoustic characteristics for three random female voices

#	Acoustic characteristics	Female 1	Female 2	Female 3	Average value	Min value	Max value
1	meanfreq	0.165508946001837	0.142056255712406	0.143658744830027	0.150407982	0.142056	0.165509
2	sd	0.0928835369116316	0.0957984262823456	0.090628260997323	0.093103408	0.090628	0.095798
3	median	0.183043922369765	0.18373123659757	0.184976167778837	0.183917109	0.183044	0.184976
4	Q25	0.0700715015321757	0.0334238741958542	0.0435081029551954	0.04900116	0.033424	0.070072
5	Q75	0.250827374872319	0.224360257326662	0.219942802669209	0.231710145	0.219943	0.250827
6	IQR	0.180755873340143	0.190936383130808	0.176434699714013	0.182708985	0.176435	0.190936
7	skew	1.70502911922022	1.8765016261382	1.59106489452459	1.724198547	1.591065	1.876502
8	kurt	5.76911536636857	6.60450859443523	5.38829754263834	5.920640501	5.388298	6.604509
9	sp.ent	0.938829422236203	0.946854262087822	0.950436378939425	0.945373354	0.938829	0.950436
10	sfm	0.601528810198165	0.65419636266412	0.675469716753195	0.64373163	0.601529	0.67547
11	mode	0.267701736465781	0.00800571837026448	0.212202097235462	0.162636517	0.008006	0.267702
12	centroid	0.165508946001837	0.142056255712406	0.143658744830027	0.150407982	0.142056	0.165509
13	meanfun	0.185606931233589	0.209917676760527	0.172374995915133	0.189299868	0.172375	0.209918
14	minfun	0.0622568093385214	0.0395061728395062	0.0344827586206897	0.045415247	0.034483	0.062257
15	maxfun	0.271186440677966	0.275862068965517	0.25	0.265682837	0.25	0.275862
16	meandom	0.227022058823529	0.494270833333333	0.791360294117647	0.504217729	0.227022	0.79136
17	mindom	0.0078125	0.0078125	0.0078125	0.0078125	0.007813	0.007813
18	maxdom	0.5546875	2.9375	3.59375	2.361979167	0.554688	3.59375
19	dfrange	0.546875	2.9296875	3.5859375	2.354166667	0.546875	3.585938
20	modindx	0.35	0.194758620689655	0.311002178649237	0.2852536	0.194759	0.35

Production rules were determined based on the results of the analysis of 3,168 recorded voice samples:

$$P = \{\alpha(?\,c_{mode}, 0.14) \wedge \beta(?\,c_{mode}, 0.059) \rightarrow male,$$
$$\alpha(?\,c_{mode}, 0.012) \wedge \alpha(?\,c_{maxdom}, 6.6) \rightarrow male,$$
$$\alpha(?\,c_{mode}, 0.012) \wedge \beta(?\,c_{maxdom}, 6.6) \rightarrow female,$$
$$\alpha(?\,c_{mode}, 0.059) \wedge \beta(?\,c_{mode}, 0.012) \rightarrow female,$$
$$\beta(?\,c_{mode}, 0.14) \wedge \beta(?\,c_{minfun}, 0.22) \rightarrow male,$$
$$\beta(?\,c_{mode}, 0.21) \wedge \alpha(?\,c_{minfun}, 0.22) \wedge \alpha(?\,c_{Q25}, 0.5) \wedge \beta(?\,c_{meanfun}, 1.6) \wedge \alpha(?\,c_{skew}, 12) \rightarrow male,$$
$$\beta(?\,c_{mode}, 0.21) \wedge \alpha(?\,c_{minfun}, 0.22) \wedge \alpha(?\,c_{Q25}, 0.5) \wedge \beta(?\,c_{meanfun}, 1.6) \wedge \beta(?\,c_{skew}, 12) \rightarrow female,$$
$$\beta(?\,c_{mode}, 0.21) \wedge \alpha(?\,c_{minfun}, 0.22) \wedge \alpha(?\,c_{Q25}, 0.5) \wedge \alpha(?\,c_{meanfun}, 1.6) \rightarrow female,$$
$$\beta(?\,c_{mode}, 0.21) \wedge \alpha(?\,c_{minfun}, 0.22) \wedge \beta(?\,c_{Q25}, 0.5) \wedge \alpha(?\,c_{median}, 0.8) \rightarrow male,$$
$$\beta(?\,c_{mode}, 0.21) \wedge \alpha(?\,c_{minfun}, 0.22) \wedge \beta(?\,c_{Q25}, 0.5) \wedge \beta(?\,c_{median}, 0.8) \rightarrow female\}$$

Next, we will build a machine learning model (Logistic Regression, SVM, Random Forest, kNN and Decision Tree) [[6], [25], [26], [47], [48], [49]], which will learn to distinguish between male and female voices based on these acoustic features. After training the model, we will test it with new voice recordings that were not used during training. We will evaluate the accuracy and reliability of our gender recognition system using metrics such as accuracy, sensitivity, specificity, etc. [[21], [28], [29], [33], [38], [46]]. The task includes the following steps:

1) Pre-processing and Collection of the sound data set, which includes male and female voice recordings.
2) Extraction of acoustic features from the voice recordings, such as frequency characteristics, spectral characteristics, amplitude, modulation, etc.
3) Splitting the dataset into a testing set and a training set.
4) Building a machine/deep learning model to classify voices by gender.
5) Model training on the training dataset.
6) Evaluation of the model using a test dataset and calculation of metrics to assess the accuracy and reliability of the gender recognition system.
7) Using the trained model to recognize gender in new voice recordings.

Successful implementation of the project involves the reliable model construction for voice gender recognition that can be applied in various areas, such as automatic recognition of voice commands, authentication systems, interactive

voice assistants, etc. [[4], [31], [32], [40], [41], [42], [53], [54], [59]]. Technical specifications for our voice gender recognition project include the following milestones and requirements:

1) Data collection and pre-processing:

- Preparation of the sound data set, which includes male and female voice recordings.
- Each voice recording must be presented in the appropriate audio format (WAV in our case).
- Data processing to remove unwanted noise, equalize volume, etc.

2) Extraction of acoustic features:

- Development of algorithms for extracting various acoustic features from voice recordings, such as frequency characteristics, spectral characteristics, amplitude, modulation, etc.
- Selection of optimal acoustic features that best reflect the differences between male and female voices.

3) Building of classification model:

- Choosing a machine/deep learning algorithm to build a gender classification model capable of recognizing gender by voice.
- Preparation of the training dataset and its division into training and validation subsets.
- Setting the parameters of the model and training the model using the training dataset.

4) Model assessment and improvement:

- Using the validation data subset to assess model accuracy and reliability.
- Analysis of results and improvement of the model by changing parameters, algorithms or adding new features.
- Using metrics such as sensitivity, accuracy, specificity, etc. to evaluate model performance.

5) Model testing and validation:

- Using the test dataset for final validation of the model and its overall performance evaluation.
- Model validation with different types of voice recordings and under different conditions to ensure its universality and stability.

6) Implementation of the gender recognition system:

- Integration of the trained model into the voice recognition system.
- Development of a user interface that will allow inputting voice data for gender recognition.
- Development of an algorithm for processing voice data before submitting it to the model.
- Implementation of a mechanism for processing the model's response and outputting the gender recognition result.

7) System testing and optimization:

- Testing the gender recognition system on real input data and evaluating its performance and reliability.
- Identification and correction of errors, and improvement of the system considering the received test results.
- Optimizing the system to achieve greater speed and performance, in particular using optimization of calculations and resources.

8) Documentation:

- Preparation of technical documentation that describes the entire development process, methodologies, used algorithms, model parameters, test results, and recommendations for the use of the system.

9) Project management:

- Planning and organizing of work on the project, setting deadlines and task priorities.
- Coordination of the development team, and communication with stakeholders and customers.
- Reporting, progress monitoring and ensuring that tasks are completed according to a schedule.

4. Experiments

At first, we upload the dataset with Kaggle [[1]] to train a system to analyse voice during gender recognition (Fig. 1-2). The dataset is called *Gender Recognition by Voice.* The dataset was created to identify voice gender based on acoustic characteristics of the voice and speech. A dataset consists of 3,168 recorded voice samples collected from female (1,584 records) and male (1,584 records) speakers. Voice samples are pre-processed based on acoustic analysis in the R environment and the *tuneR* and *seewave* packages, whose analysed frequency range is from 0 Hz to 280 Hz (the frequency range of a human voice). Within the scope of the study, no pre-processing, noise processing and balance between the samples of men (1584) and women (1584) was set in the analysis of the used data set (recognition of

gender by voice). Because the author of the dataset, Kory Becker, has already done it before us (License CC BY-NC-SA 4.0). But in living conditions, for example, when teaching people with special needs, there will be no gender balance of students, there will be informational noise, etc. Due to the inability to collect data due to ethical considerations related to confidentiality, consent, and bias, it was decided to use an open-source dataset for the study, with data collected from the following resources:

- VoxForge Speech Corpus;
- Festvox CMU_ARCTIC Speech Database at Carnegie Mellon University;
- The Harvard-Haskins Database of Regularly-Timed Speech,
- Telecommunications & Signal Processing Laboratory (TSP) Speech Database at McGill University.

The data set used is freely available for conducting various studies, including for analysis of the comparability of various types of data.

# meanfreq	# sd	# median	# Q25	# Q75	# IQR	# skew	# kurt	# sp.ent
0.04 – 0.25	0.02 – 0.12	0.01 – 0.26	0 – 0.25	0.04 – 0.27	0.01 – 0.25	0.14 – 34.7	2.07 – 1.31k	0.74
0.0597809849598081	0.0642412677031359	0.032026913372582	0.0150714886459209	0.0901934398654331	0.0751219512195122	12.8634618371626	274.402905502067	0.893369416700807
0.066008740387572	0.0673100287952527	0.040228734810579	0.0194138670478914	0.0926661901358113	0.0732523230879199	22.4232853628204	634.613854542068	0.892193242265734
0.0773155026958227	0.0838294209445061	0.0367184586699814	0.00870105655686762	0.131908017402113	0.123206960845246	30.7571545800584	1024.927704721	0.846389091878782
0.151228091724635	0.0721105872627985	0.158011187072716	0.0965817277812306	0.207955251709136	0.111373523927906	1.23283127554372	4.17729621025317	0.963322461535984
0.135120387296677	0.0791461004935869	0.124656228727025	0.0787202178352621	0.206044928522805	0.127324710687543	1.10117366551545	4.3337131547091	0.971955076212905
0.132786407306188	0.0795568659729794	0.119089848308051	0.0679579929988331	0.209591598599767	0.141633605600933	1.93256243208397	8.30889503672364	0.963181346316771
0.150762330177465	0.0744632052034133	0.160106382978723	0.0928989361702128	0.205718085106383	0.11281914893617	1.53064323137138	5.98749765804991	0.967573069931702
0.160514332166817	0.0767668848761485	0.144336775218427	0.110532168387609	0.231961874503574	0.121429706115965	1.39715636508704	4.76661069503212	0.959254569785851
0.142239416807037	0.0780184624458836	0.138587443946188	0.0882062780269058	0.208587443946188	0.120381165919283	1.09974615121553	4.07028370586557	0.970722932180341
0.134328777309825	0.0803500276143771	0.121451346893898	0.0755799890049478	0.201957119296317	0.126377130291369	1.19036844433836	4.78730983303407	0.975246142524797
0.157020511056795	0.0719429306811688	0.168160152526215	0.101429933269781	0.2167397521449	0.115309818875119	0.97944227542161	3.97422262552131	0.965249139287411
0.138550521389033	0.0770539929982884	0.127526595744681	0.087313829787234	0.202739361702128	0.115425531914894	1.62676986679461	6.29136507870849	0.966003775055692
0.137342741880723	0.0808767069335911	0.124262508122157	0.0831448992852502	0.209226770630279	0.126081871345029	1.37872820350982	5.00895162500433	0.963513527788451
0.181225459373767	0.0600420545194889	0.190953206239168	0.128838821490468	0.229532062391681	0.100693240901213	1.36943045758527	5.47559978801453	0.937445760468646
0.183115280687352	0.0669823503730641	0.191232528589581	0.129148665819568	0.240152477763659	0.111003811944091	3.56810402353109	35.3847484526494	0.940332622910243

```
1  meanfreq,"sd","median","Q25","Q75","IQR","skew","kurt","sp.ent","sfm","mode","centroid","meanfun","minfun","maxfun","meandom","mindom","maxdom","dfrange","modindx","label"
2  0.0597809849598081,0.0642412677031359,0.032026913372582,0.0150714886459209,0.0901934398654331,0.0751219512195122,12.8634618371626,274.402905502067,0.893369416700807,0.491917766
3  0.066008740387572,0.0673100287952527,0.040228734810579,0.0194138670478914,0.0926661901358113,0.0732523230879199,22.4232853628204,634.613854542068,0.892193242265734,0.5137238425
4  0.0773155026958227,0.0838294209445061,0.0367184586699814,0.00870105655686762,0.131908017402113,0.123206960845246,30.7571545800584,1024.927704721,0.846389091878782,0.478904979110
5  0.151228091724635,0.0721105872627985,0.158011187072716,0.0965817277812306,0.207955251709136,0.111373523927906,1.23283127554372,4.17729621025317,0.963322461535984,0.727231798861
6  0.135120387296677,0.0791461004935869,0.124656228727025,0.0787202178352621,0.206044928522805,0.127324710687543,1.10117366551545,4.3337131547091,0.971955076212905,0.783568057553
7  0.132786407306188,0.0795568659729794,0.119089848308051,0.0679579929988331,0.209591598599767,0.141633605600933,1.93256243208397,8.30889503672364,0.963181346316771,0.738307004485
8  0.150762330177465,0.0744632052034133,0.160106382978723,0.0928989361702128,0.205718085106383,0.112819148936171,1.53064323137138,5.98749765804991,0.967573069931702,0.762637670276
9  0.160514332166817,0.0767668848761485,0.144336775218427,0.110532168387609,0.231961874503574,0.121429706115965,1.39715636508704,4.76661069503212,0.959254569785851,0.719857907046
10 0.142239416807037,0.0780184624458836,0.138587443946188,0.0882062780269058,0.208587443946188,0.120381165919283,1.09974615121553,4.07028370586557,0.970722932180341,0.770992054490
11 0.134328777309825,0.0803500276143771,0.121451346893898,0.0755799890049478,0.201957119296317,0.126377130291369,1.19036844433836,4.78730983303407,0.975246142524797,0.804505253038
12 0.157020511056795,0.0719429306811688,0.168160152526215,0.101429933269781,0.2167397521449,0.115309818875119,0.97944227542161,3.97422262552131,0.965249139287411,0.736928779287570
13 0.138550521389033,0.0770539929982884,0.127526595744681,0.087313829787234,0.202739361702128,0.115425531914894,1.62676986679461,6.29136507870849,0.966003775055692,0.752040210269243
14 0.137342741880723,0.0808767069335911,0.124262508122157,0.0831448992852502,0.209226770630279,0.126081871345029,1.37872820350982,5.00895162500433,0.963513527788451,0.736149955067
15 0.181225459373767,0.0600420545194889,0.190953206239168,0.128838821490468,0.229532062391681,0.100693240901213,1.36943045758527,5.47559978801453,0.937445760468646,0.537079994941
16 0.183115280687352,0.0669823503730641,0.191232528589581,0.129148665819568,0.240152477763659,0.111003811944091,3.56810402353109,35.3847484526494,0.940332622910243,0.571394201518
17 0.174272105181833,0.0694110453235828,0.190874106652007,0.115601979109401,0.228279274326553,0.112677295217152,4.48503835015822,61.7649083141473,0.950972024663856,0.635199178449
18 0.190846297652897,0.0657902823878029,0.207950987066031,0.132280462899932,0.244356705241661,0.112076242341729,1.56230368302863,7.83434988641618,0.938546013440116,0.538809584118
19 0.171246971349986,0.074871569699177,0.152806652806653,0.122390852390852,0.243617463617464,0.121226611226611,3.20716982695899,25.765564944839,0.936953536867586,0.586419514172829,
20 0.168345946168349,0.074120817065231,0.145618471687741,0.115755909840572,0.239824079164376,0.124068169323804,2.70433472872787,18.4847025968934,0.934523379604557,0.559742318266040
21 0.173631376053858,0.073351512008514,0.153568557391693,0.12368021438142,0.244234033050469,0.120553818669049,2.80497488480386,20.8575431560588,0.930916892819773,0.518268556530761,
22 0.172754287202847,0.0769025830140396,0.177736289381564,0.120070011668611,0.245367561260210,0.125297549591599,2.96776509333623,20.0781146980326,0.925539326979812,0.523081122573215
23 0.181015393833521,0.0743693038331353,0.169298813376483,0.128673139158576,0.254174757281553,0.125501618122977,2.58732453355436,12.2814319226622,0.915284119626616,0.475317470815
24 0.163535853470868,0.0724488983130930,0.145542521994135,0.113929618768328,0.227448680351906,0.113519061583578,3.58764999004366,28.6537805052841,0.927014678290766,0.542422403541930
25 0.170212797508261,0.0751049460722898,0.1460529986053,0.123988842398884,0.250125523012552,0.126136680613668,2.81679291835158,13.7645822009404,0.913832253608624,0.487965715691965,
26 0.160422476518119,0.0766147644560479,0.144823805794832,0.120924040720439,0.237243539545811,0.116319498825372,6.25320825198848,85.4919256914162,0.933029647028206,0.567423613537
27 0.164699838046527,0.0753617735918,0.147018469656992,0.118698328935796,0.240474934036939,0.121776605101143,4.20860840517055,43.6818854251652,0.940668790234324,0.604020485106291,
28 0.169578801414252,0.0756345139572698,0.186467576791809,0.116706484641638,0.238549488054608,0.121843003412969,4.26992267782145,45.895247817628,0.929498042215949,0.543708544075525
29 0.169021003926254,0.0717781520946695,0.143168482816429,0.12580050293378,0.248315171835708,0.122514668901928,3.07927339517432,14.3402988428128,0.902275256056735,0.477745723075478
30 0.167340238692684,0.0728412302318368,0.141739130434783,0.122173913043478,0.24,0.117826086956522,2.19212624568844,8.1524097848739,0.913763493306731,0.539479089755153,0.134782608
31 0.180527821446414,0.0708669980737599,0.142385321100917,0.129541284404367,0.252477064220184,0.122935779816514,2.79996894420966,12.1903605225796,0.853114893695345,0.313426449697903
32 0.153039047583376,0.0740311299735897,0.158064516129032,0.0927311827956989,0.214365591397849,0.121634408602151,0.885701647564837,3.52398185555844,0.973217841619629,0.807551708403
33 0.166659498779566,0.075961167558551,0.142781456953642,0.122075055187638,0.245077262693157,0.123002207505519,3.28184796412453,16.6194816902798,0.915784869319231,0.560406695288123
34 0.175659428789724,0.0716515330560574,0.144191616766467,0.131057884231537,0.256526946107784,0.125469061876248,3.73648728388072,21.6666687322304,0.876748955245590,0.403909559814853,
35 0.174825733261765,0.0715328311457178,0.146470588235294,0.123529411764706,0.247058823529412,0.123529411764706,2.57673204167683,10.1101248635807,0.875391518945669,0.436706209916
36 0.174064613695703,0.073121843993196,0.140183486238532,0.127706422018349,0.252844036697248,0.125137614678899,3.26303975514635,16.4240546533024,0.875611110757132,0.414838884375250
```

Fig. 1. Dataset example

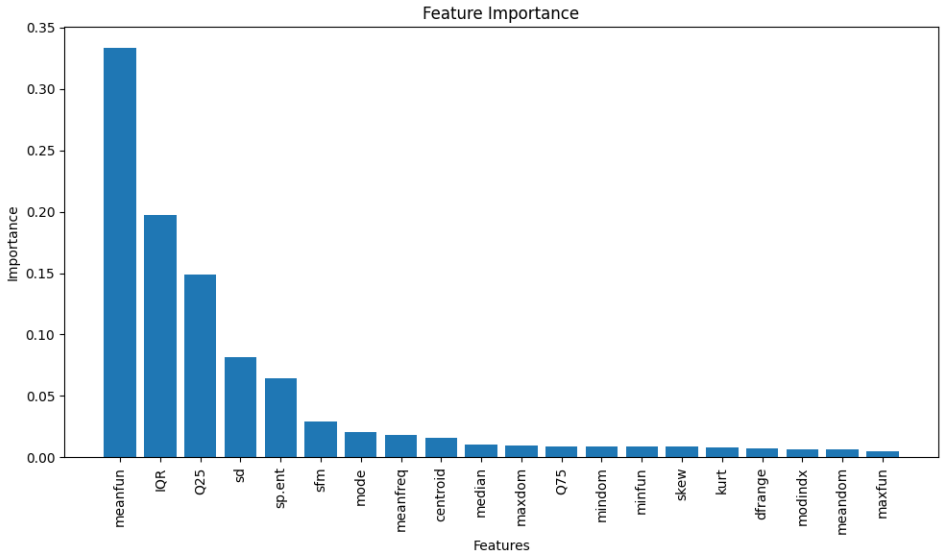

Fig. 2. Visualization of the model feature importance graph

The Python programming language and the *pandas* library are used for data analysis and processing. I will work in the Jupyter Notebook environment.

1) Import *pandas*, read the input file, and display the size of the dataset (Fig. 3):

```
import pandas as pd
[1] ✓ 18.1s

data = pd.read_csv('voice.csv')
[2] ✓ 0.4s

(3168, 21))
```

Fig. 3. Import of pandas, file reading and dataset size checking

2) Display the first five entries in the dataset and display attribute data (Fig. 4):

	meanfreq	sd	median	Q25	Q75	IQR	skew
0	0.059781	0.064241	0.032027	0.015071	0.090193	0.075122	12.863462
1	0.066009	0.067310	0.040229	0.019414	0.092666	0.073252	22.423285
2	0.077316	0.083829	0.036718	0.008701	0.131908	0.123207	30.757155
3	0.151228	0.072111	0.158011	0.096582	0.207955	0.111374	1.232831
4	0.135120	0.079146	0.124656	0.078720	0.206045	0.127325	1.101174

	kurt	sp.ent	sfm	...	centroid	meanfun	minfun	\
0	274.402906	0.893369	0.491918	...	0.059781	0.084279	0.015702	
1	634.613855	0.892193	0.513724	...	0.066009	0.107937	0.015826	
2	1024.927705	0.846389	0.478905	...	0.077316	0.098706	0.015656	
3	4.177296	0.963322	0.727232	...	0.151228	0.088965	0.017798	
4	4.333713	0.971955	0.783568	...	0.135120	0.106398	0.016931	

	maxfun	meandom	mindom	maxdom	dfrange	modindx	label
0	0.275862	0.007812	0.007812	0.007812	0.000000	0.000000	male
1	0.250000	0.009014	0.007812	0.054688	0.046875	0.052632	male
2	0.271186	0.007990	0.007812	0.015625	0.007812	0.046512	male
3	0.250000	0.201497	0.007812	0.562500	0.554688	0.247119	male
4	0.266667	0.712812	0.007812	5.484375	5.476562	0.208274	male

[5 rows x 21 columns])

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3168 entries, 0 to 3167
Data columns (total 21 columns):
 #   Column    Non-Null Count  Dtype
---  ------    --------------  -----
 0   meanfreq  3168 non-null   float64
 1   sd        3168 non-null   float64
 2   median    3168 non-null   float64
 3   Q25       3168 non-null   float64
 4   Q75       3168 non-null   float64
 5   IQR       3168 non-null   float64
 6   skew      3168 non-null   float64
 7   kurt      3168 non-null   float64
 8   sp.ent    3168 non-null   float64
 9   sfm       3168 non-null   float64
 10  mode      3168 non-null   float64
 11  centroid  3168 non-null   float64
 12  meanfun   3168 non-null   float64
 13  minfun    3168 non-null   float64
 14  maxfun    3168 non-null   float64
 15  meandom   3168 non-null   float64
 16  mindom    3168 non-null   float64
 17  maxdom    3168 non-null   float64
 18  dfrange   3168 non-null   float64
 19  modindx   3168 non-null   float64
 20  label     3168 non-null   object
dtypes: float64(20), object(1)
memory usage: 519.9+ KB
```

Fig. 4. First 5 dataset records and attribute information

3) Create and visualize the correlation matrix of attributes (Fig. 5).

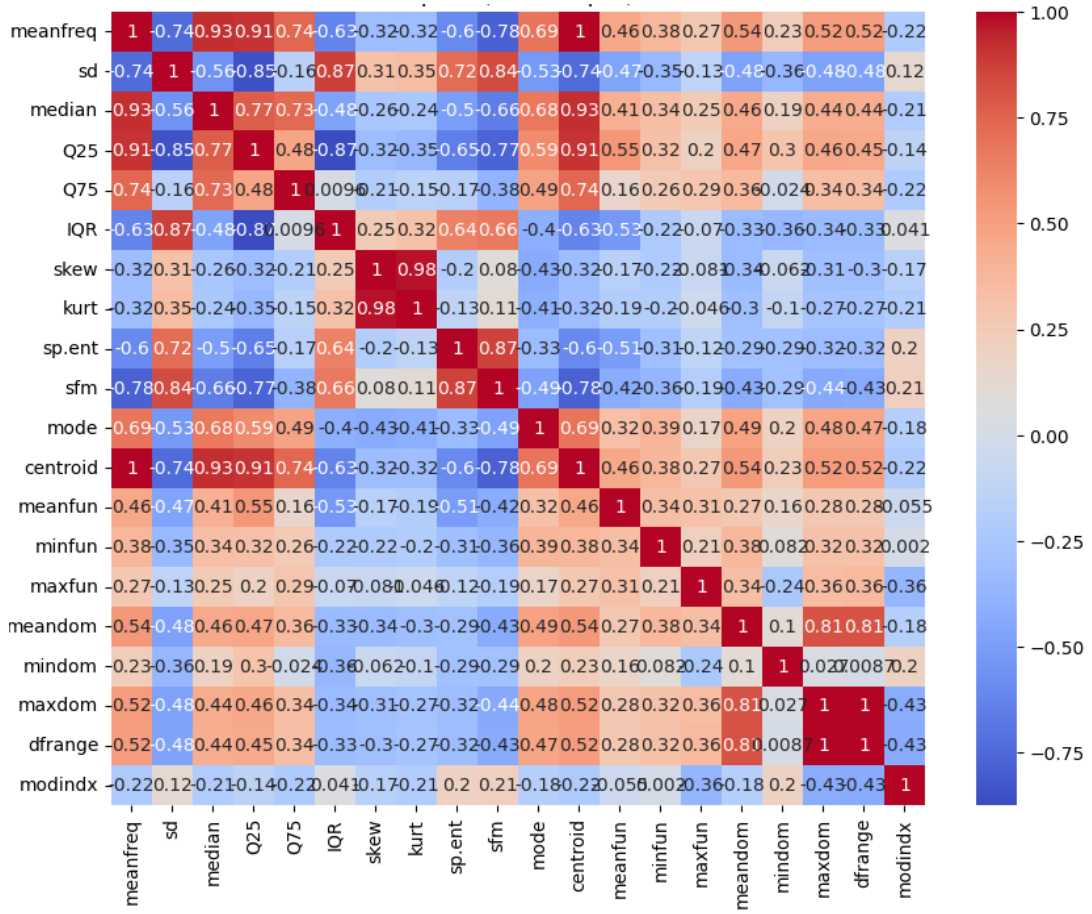

Fig. 5. Correlation matrix of attributes

Since the program will work with input files and not with a dataset, it is necessary to implement a voice processing mechanism and obtain the necessary numerical parameters that will allow obtaining a result for each case.

Voice processing algorithm

Stage 1. Calculation of sound parameters based on statistical analysis of the function [modindx, meanfreq, dfrange, sd, maxdom, median, mindom, q25, meandom, q75, maxfun, iqr, minfun, skew, meanfun, kurt, peakf, sp_ent, centroid, sfm, mode].
Stage 2. Analysis of spectrum entropy.
Step 2.1. Performing a Fourier transform on audio data.
Step 2.2. Calculation of the normalized spectrum.
Step 2.3. Calculation of spectrum entropy.
Stage 3. Conversion of audio data into a spectrogram and spectral uniformity calculation.
Stage 4. Calculation of centroid frequency.
Stage 5. Peak frequency analysis.
Step 5.1. Audio data conversion into a spectrogram.
Step 5.1. Finding the index of the maximum spectrum value.
Step 5.1. Obtaining the peak frequency from the corresponding bin.
Stage 6. The modulation index analysis.
Step 6.1. Calculation of signal amplitude.
Step 6.1. Calculation of the difference between adjacent amplitudes.
Step 6.1. Calculation of the modulation index.

To implement a full-fledged voice analysis project, it is necessary to define several files:

1)program.py – the file where code execution begins – gender identification of a voice in the audio file.
2)proccess.py – the file for actions with the audio file and its analysis
3)train.py – the file for training the model

Definition of the proccess.py file:

Step 1. Average accuracy.
Step 2. Cut out blank spaces at the beginning and the end.
Step 3. Add silence at the beginning and at the end that is 'seconds' long (a variable of the float type).
Step 4. Record a speaker using a microphone and save the file to 'path'.

This code contains functions to record voice from the microphone and get voice parameters using the *librosa* library. Main functions:

1. **array_normalize(data)**: This function normalizes the data array by averaging the values' accuracy.
2. **trim_spaces(data)**: The function trims blank spaces (silence) at the beginning and at the end of the data array.
3. **add_silence(data, seconds)**: The function adds silence of a set length at the beginning and the end of the data array.
4. **record_voice()**: The function records a speaker's voice from the microphone. The *pyaudio* library is used to interact with the audio interface. The recording continues until the set level of silence is reached. The function returns the result of the recording as an array.
5. **record_to_file(path)**: The function records voice from the microphone and saves it in the defined file using the *wave* library.

Finally, the train.py file searches for the main parameters of the voice for their further analysis:

Step 1. Performing a Fourier transform on audio data.
Step 2. Calculation of the normalized spectrum.
Step 3. Calculation of spectrum entropy.
Step 4. Conversion of audio data into a spectrogram and spectral uniformity calculation.
Step 5. Calculation of centroid frequency.
Step 6. Finding the maximum spectrum value index.
Step 7. Obtaining the peak frequency from the corresponding bin.
Step 8. Calculation of the difference between adjacent amplitudes.
Step 9. Calculation of the modulation index.
Step 10. Calculation of sound parameters [modindx, meanfreq, dfrange, sd, maxdom, median, mindom, q25, meandom, q75, maxfun, iqr, minfun, skew, meanfun, kurt, peakf, sp_ent, centroid, sfm, mode]

5. Results. Basic Material

Let us consider the execution of our code and demonstrate its operation. First, an ordinary male voice is uploaded in the WAV format. To open this file, we use the code from the previous work. The program starts training the model

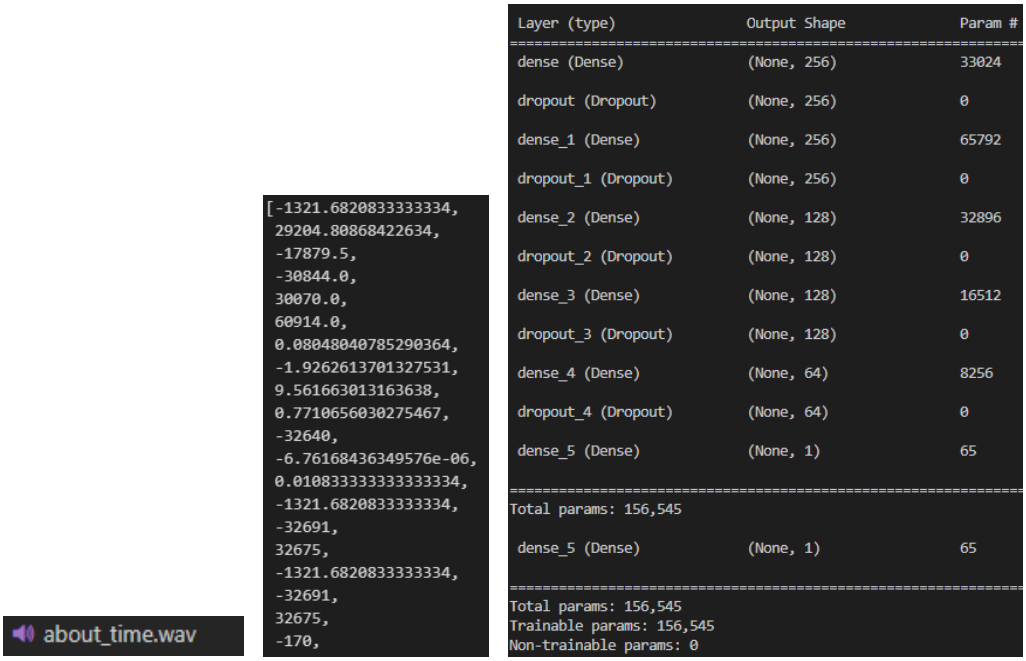

Fig. 6. Audio file, performance results and model training

(Fig. 6). After that, the input file is processed and the result is output. The file from previous works and this file contains a male voice (male 96.36% and female 3.64%). Let us try to test a model, but already with the voice of the first co-author of this publication. The program can listen to the speaker's voice. We enter the input command: python test.py. The program issues the command "You can talk" when it is necessary to start speaking and immediately after the end of the speech it starts analyzing and outputs the following result: male 97.37% and female 2.63%. Let us try to do the same, but this time with a female voice, and for this there is the specially defined file in the folder of our program. The execution begins with model training (Fig. 7).

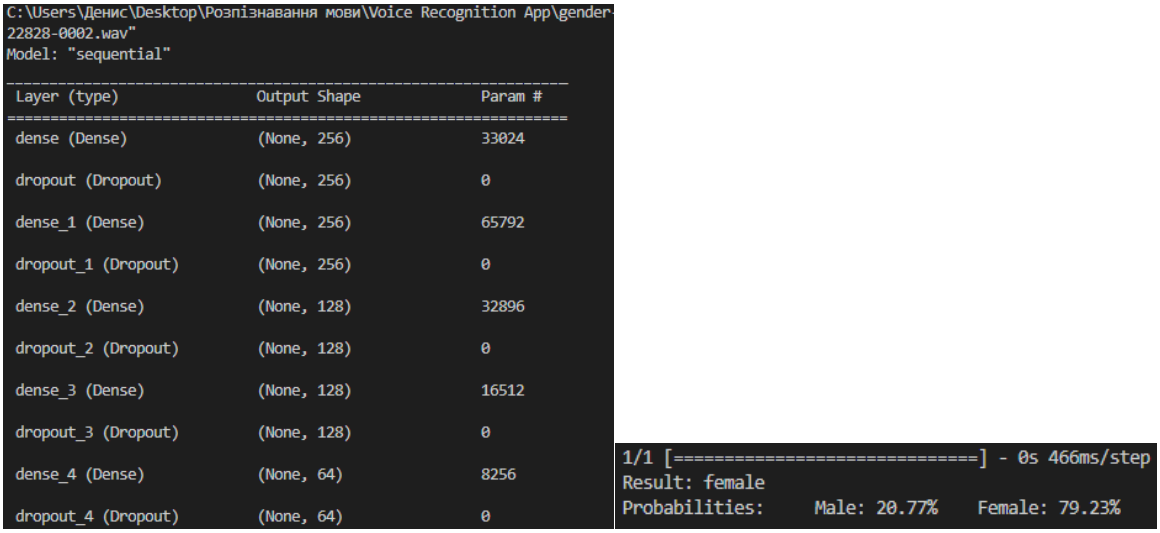

Fig. 7. Model training and program execution results

The model worked correctly, although it gave a relatively high percentage of the probability of a male voice. This code performs voice gender recognition using a support vector machine, the decision tree and logistic regression. It calculates accuracy, precision, and recall sensitivity for each algorithm.

1. Accuracy: It is the ratio of the correct predictions' number to the total number of observations. It measures the overall accuracy of the model and is expressed as a percentage. A higher precision indicates a more accurate model. Accuracy = (TP + TN) / (TP + TN + FP + FN), where: TN (True Negative) is a correctly predicted negative classes' number; TP (True Positive) is a correctly predicted positive classes' number; FN (False Negative) is an incorrectly predicted negative classes' number; FP (False Positive) is an incorrectly predicted positive classes' number.
2. Precision: It is the ratio of a correctly predicted positive class number to a total number of positive predictions. It measures how accurate the predictions of the positive class are. Precision formula: Precision = TP / (TP + FP).
3. Recall: It is the ratio of a correctly predicted positive class's number to the total number of valid positive classes. It measures how well the model detects positive predictions. Recall formula: Recall = TP / (TP + FN).

Algorithm of voice classification by gender based on machine learning

Step 1. Importing all the necessary libraries and uploading the dataset.
Step 2. To evaluate the models, it is necessary to separate the target variable from other features.
Step 3. Splitting data into testing and training datasets.
Step 4. Next, we standardize the attributes. Attribute standardization is the bringing attribute values process to a standard scale or range. It is used to ensure the same scale for all attributes, which facilitates data comparison and analysis. Attribute standardization is usually done by subtracting the attribute mean from each attribute value and dividing by a standard deviation. This transformation ensures that the attribute mean is zero and the standard deviation is one. Standardization formula: x' = (x - mean) / std, where: x is the initial attribute value; x' is a standardized attribute value; mean is the mean attribute value; std is a standard deviation of an attribute. Then, we move to the next step is building different methods of machine learning based on SVM, logistic regression and decision trees.
Step 5. Implementation of logistic regression as logreg_predictions = logreg.predict(X_test).
Step 6. Implementation of a decision tree as DecisionTreeClassifier().
Step 7. Implementation of an SVM as svm = SVC().
Step 8. We evaluate f each model separately for performance for male and female voices.
Step 9. We display the results on a screen (Fig. 8):

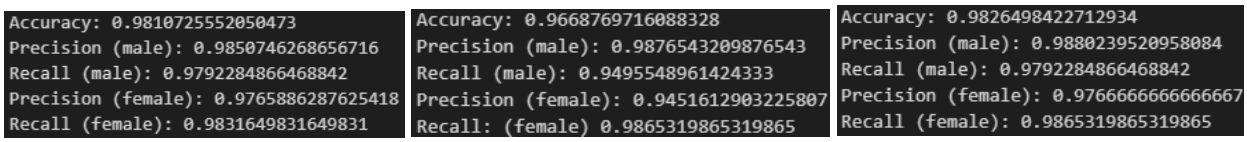

Fig. 8. The results of the logistic regression analysis, the decision tree and the support vector machine

When we compare the three models, we can see that a support vector machine showed the best accuracy – 98.2%. The other two indicators – precision and recall – are also the highest in this method. Logistic regression comes second, and a decision tree has the worst results.

Next, we implement an experimental comparison of different classification algorithms for voice recognition. This code performs cross-validation for various voice classification algorithms: decision tree, SVM, logistic regression, and random forest. It uses cross-validation with 5 sections (cv=5) and outputs the average accuracy and standard deviation for each algorithm. This experiment allows comparing the performance of different voice classification algorithms on the same dataset and determining which shows the best results. The results help decide which algorithm is the most suitable for a specific voice recognition task, depending on its predictions' accuracy and stability.

Stage 1. Importing all the necessary libraries, uploading the dataset, dividing the target variable from other attributes, and standardizing the attributes.

Stage 2. Defining 4 classification models – decision trees, logistic regression, SVM, and random forest.

Stage 3. Cross-validation of each method.

Stage 4. Comparing algorithms using cross-validation. Deriving average values of accuracy and standard deviation for each algorithm and checking the comparison results:

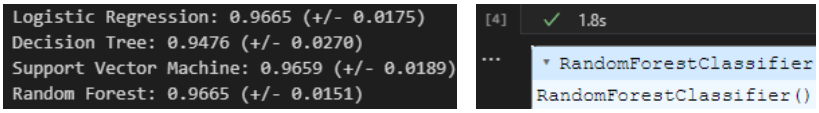

Fig. 9. Result of comparing algorithms using cross-validation and result of building the RandomForestClassifier model

According to the results, the decision tree has the worst accuracy, while the other three methods are more accurate. Considering the indicator together with "plus-minus", the support vector machine is potentially the best again, although it took third place initially.

Analysing the obtained results (1) with the results of https://www.primaryobjects.com/2016/06/22/identifying-the-gender-of-a-voice-using-machine-learning/ (2) [1] and https://www.kaggle.com/datasets/primaryobjects/voicegender (3) [2] we will build a comparative table (Table 3).

Table 3. Comparative analysis of the training results in accuracy of the Gender classification models

#	Method	Accuracy		
		(1)	(2)	(3)
1	Logistic Regression	96% / 98%	71% / 72%	97% / 98%
2	Decision Tree	94% / 95%	-	-
3	SVM	96% / 97%	96% / 85%	100% / 99%
4	Random Forest	96% / 97%	100% / 87%	100% / 98%
5	CART	96% / 97%	81% / 78%	96% / 97%
6	Baseline (always predict male)	-	61% /59%	50% / 50%
7	Boosted Tree	94% / 95%	91% / 84%	-
8	XGBoost	97% / 98%	100% / 87%	100% / 99%
9	kNN	98% / 99%	89% / 100%	-
10	neural networks	98% / 99%	-	-

After reducing the frequency ranges within the human voice to 0-280 Hz with a sound threshold of 15%, the accuracy increases to 97%/100%.

Next, let us determine a feature's importance in the voice recognition task. In this work, we will use the Random Forest model to determine a feature's importance in voice recognition tasks. It calculates each feature's importance using the feature_importances_ attribute of the RandomForestClassifier model. Next, the code creates a graph that visualizes a feature's importance. A graph will help us understand which features have the greatest impact on voice recognition and can be used to improve the model.

Stage 1. Definition and import of libraries. Uploading the dataset and dividing the target feature from other features.

Stage 2. Building the random forest model RandomForestClassifier (Fig. 9).

Stage 3. Defining the 'importance' variable, which will be responsible for a feature's importance.

Stage 4. Creating the data frame for further visualization, which will contain two fields – Feature (feature name) and Importance (feature importance).

Stage 5. Configuring and visualizing a graph using the *matplotlib.pyplot* library.

As it can be seen, a feature that affects the model execution result most is meanfun (the mean value of the fundamental frequency measured indirectly by the acoustic signal, almost 35%), IQR (interquartile range, 20%) and Q25 (first quartile, 15%). These 3 features determine the result of the model by almost 70%. The lowest indicators are for such features as maxfun, meandom, modindx, and others.

Let us classify a voice using neural networks. In this work, a neural network using the scikit-learn library will be used for voice classification. The dataset is divided into testing and training datasets, and the features are normalized using StandardScaler. A neural network model will be built with two hidden layers, trained on training data, and then used to make predictions on testing data. Finally, the accuracy metric will be used to evaluate the model performance.

Stage 1. Definition and import of libraries, uploading of the dataset and dividing the features into the target feature and all other features.

Stage 2. Data division into a training and a test dataset.

Stage 3. Building a growing tree method model (Fig. 9).

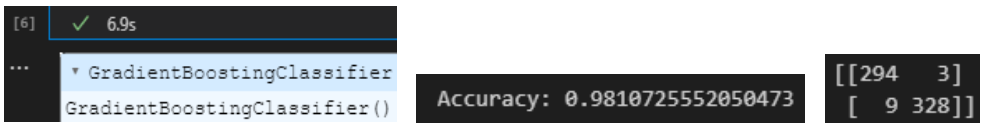

Fig. 10. Result of building the model, the growing tree method model accuracy, the confusion matrix

Stage 4. Calculation of the y_pred variable is the prediction of our label – 'Male' or 'Female':

Stage 5. Model performance evaluation and display of the result (Fig. 10).

Stage 6. Building the confusion matrix (Fig.10). A confusion matrix provides information on a correctly and incorrectly classified instances' number for each class. It allows you to evaluate the types of errors, such as false positives and false negatives, and understand the model performance.

Let us try to interpret the values of this matrix:

- In the top row of the first column, there are 294. It means that 294 instances were correctly classified as "female voices" (true negatives).
- In the top row of the second column, there are 3. It means that 3 instances were classified incorrectly as "male" voices (false positives).
- In the bottom row of the first column, there are 9. It means that 9 instances were classified incorrectly as "female" voices (false negatives).
- In the bottom row of the second column, there are 328. It means that 328 instances were correctly classified as "male voices" (true positives).

The confusion matrix allows an understanding of how the model classifies the data and what types of errors it makes. Based on these values, the following conclusions can be made:

- The model classified 294 female voices (true negatives) and 328 male voices (true positives) correctly.
- The model made 3 incorrect classifications, where a female voice was classified as male (false positives).
- The model also made 9 incorrect classifications, where a male voice was classified as female (false negatives).

As for accuracy, the model is 98.1% accurate, which, compared to the methods described in previous works, ranks second together with logistic regression.

Let us use the kNN algorithm for voice classification into male and female. A dataset is divided into testing and training datasets, and the features are normalized using StandardScaler. The kNN model is built using the nearest 5 neighbours, is trained on training data, and then is used on test data for prediction. Next, the accuracy metric is used to evaluate the model performance. Finally, the confusion matrix is calculated and displayed, and a classification report is generated, detailing the accuracy, recall rates, and the F-measure for each class. So, we define and import the libraries:

Stage 1. Uploading of the dataset and dividing the features into the target feature and all other features.

Stage 2. Dividing data into the training dataset and a test dataset.

Stage 3. Feature normalization.

Stage 4. Building the kNN model (Fig. 11).

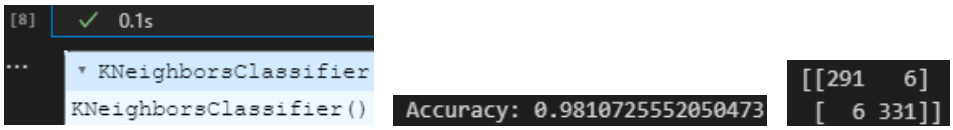

Fig. 11. Result of building the model, the kNN model accuracy, the confusion matrix

Stage 5. Calculation of the y_pred variable is the prediction of our label – 'Male' or 'Female':
Stage 6. Model performance evaluation and display of the result (Fig. 10).
Stage 7. Building the confusion matrix (Fig. 10).

- Let us try to interpret the values of this matrix:
- In the top row of the first column, there are 291. It means that 291 instances were correctly classified as "female voices" (true negatives).
- In the top row of the second column, there are 6. It means that 6 instances were classified incorrectly as "male" voices (false positives).
- In the bottom row of the first column, there are 6. It means that 6 instances were classified incorrectly as "female" voices (false negatives).
- In the bottom row of the second column, there are 331. It means that 331 instances were correctly classified as "male voices" (true positives).

Based on these values, the following conclusions can be made:

- The model classified 294 female voices (true negatives) and 328 male voices (true positives) correctly.
- Comparing the model with a growing tree method, this model shows better results in recognizing female voices, but worse results in recognizing male voices.

Let us create classification reports. A classification report provides detailed information on F1-score, recall, and precision for each class (female and male), as overall performance for an entire model. Precision reflects the percentage of correctly classified instances relative to all positive predictions. In this case, the accuracy for both classes female and male is 0.98, which means that 98% of the predictions for each class are correct. Recall reflects the percentage of correctly classified instances relative to all instances of the given class in a test set. In this case, the recall rate for both classes female and male is also 0.98, which means that the model identifies 98% of instances of each class correctly. F-measure (F1-score) is a harmonic mean between accuracy and recall. It considers both accuracy and recall. The F-measure value is 0.98 for both classes, which confirms the high performance of a model in voice recognition. The overall accuracy of the model is 0.98, which means that 98% of instances in a test dataset were recognized correctly. Analysing the classification report results, it can be concluded that our model based on a k-nearest neighbour algorithm is very effective in recognizing voices by gender. It shows high precision, recall, and F-measure for both classes. Such results indicate that the model can correctly classify a voice as male or female with high reliability. With these results, we can compete with similar voice gender recognition systems.

Let us identify differences between male and female voices using visualization and attribute analysis. In this work, a dataset is first loaded and then divided into male and female voices. Next, the distribution of each attribute in male and female voices is visualized using histograms. Each histogram is represented by a separate graph, where male voices are marked in blue and female voices in red. Next, a heat map of the correlation matrix is built to display dependencies between the attributes of the dataset. This graph allows you to identify relationships between different attributes and establish which characteristics can influence the distinction between male and female voices.

Stage 1. Definition and import of libraries, uploading of the dataset and division of attributes into the target attribute and other attributes.

Stage 2. Dataset division into female and male voices.

Stage 3. Visualization of each variable and the results are given below (only a few graphs are given in Fig. 12-14 so as not to include all the graphs in this article).

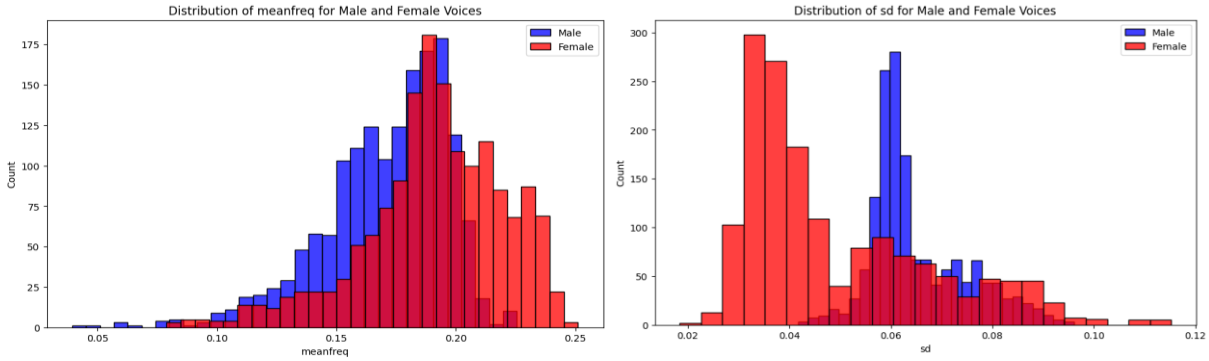

Fig. 12. Attribute meanfreq and sd Q75 for male and female voices

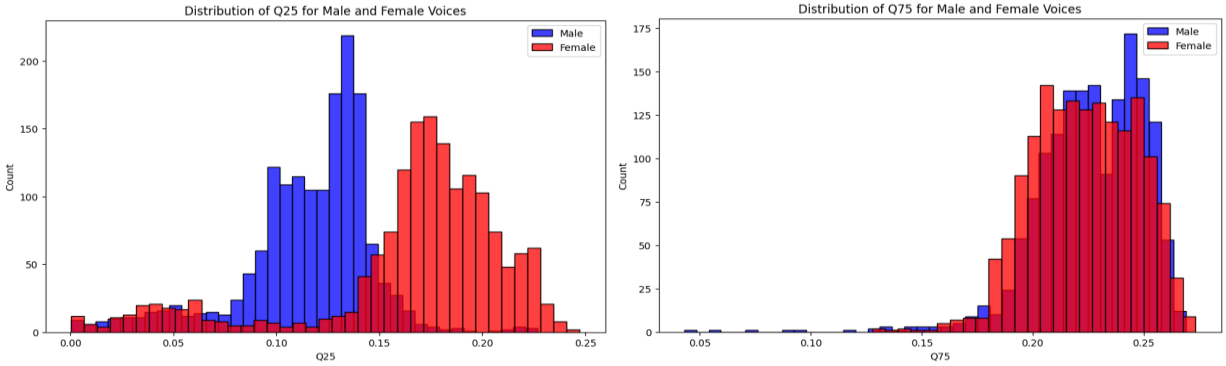

Fig. 13. Attribute Q25 and Q75 for male and female voices

Meanfun and IQR were the features that up to 70% determine whether the voice is male or female. The mean fundamental frequency is a voice gender indicator, with a threshold of 140 Hz separating female and male ranges. Below are graphs for the data on these two attributes.

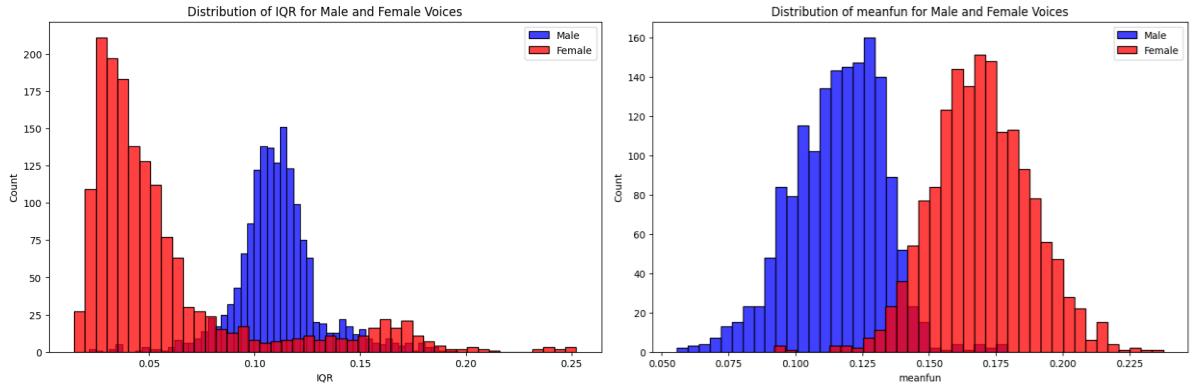

Fig. 14. IQR and meanfun for female and male voices

It is obvious that, unlike previous attribute graphs, meanfun and IQR almost clearly define male and female voices.

5. Discussion and Conclusions

In this project, we have developed a voice-based gender recognition system that can have many applications and benefit people in many ways. With the help of this system, users can transmit textual information by voice, which can then be processed and used for various purposes for example in modern education for students and teachers with vision problems or distance learning. One of the main advantages of this project is convenience. Instead of typing text manually, users are allowed to speak and their voice will be converted into text. This is especially useful for people who have limitations in writing or writing speed, and for situations where hands are busy or unavailable. It can help people with visual impairments or other physical disabilities receive information by simply speaking. This makes information more accessible and helps improve their lives. In addition, our project can find applications in automated systems that require processing of a large amount of text.

The following steps were made within our voice gender recognition project:

1. Data uploading and preparation: We have developed a function to load data and preprocess it before further processing by the model.
2. Data division into training, validation, and test datasets: Splitting one dataset into separate sets allowed us to train, configure, and evaluate the model effectively.
3. Model creation: A model architecture was developed using the library TensorFlow and Keras.
4. Model training: The model was trained using the training dataset, validation, and callbacks, such as TensorBoard and EarlyStopping.
5. Model performance evaluation: A model performance was evaluated using a test dataset and by measuring loss and accuracy.
6. Saving the trained model: The trained model was saved in the file system for further use and deployment.

In our voice gender recognition project, we used a model built based on the neural network using the TensorFlow library and Keras.

The basic model we created is a typical neural network for text classification. It consists of the input layer, hidden layers, and the output layer. For text processing, we use a pre-trained word vector space such as Word2Vec or GloVe, which provides the words' representation as numeric vectors. This allows the model to work with vector values instead of text itself. We also used such techniques as dropout to prevent model overtraining, such activation functions as ReLU (Rectified Linear Unit) for non-linearity, and a softmax function in the last layer to obtain class probabilities.

To train a model, we used the Adam optimizer, which is a popular gradient descent optimization method, and the "sparse categorical cross-entropy" loss function, since we are dealing with multi-class classification. After training the model, we saved it to a file for further use and evaluation of new data. The application of neural networks in our project allowed us to build a powerful model that can recognize a speaker's gender by voice with high accuracy. Additional techniques such as regularization and optimization can be used to improve model performance and prevent overtraining.

The following are the prospects for further research:

1. A model improvement. In the future, different approaches can be considered to improve the model, such as changing the architecture, adding layers, or using other optimization techniques.
2. Additional functionality. Additional features such as voice-based language recognition, emotion analysis or the identification of other attributes can also be considered.
3. Expansion of the dataset. More diverse data can be collected to improve voice recognition and extend model functionality.
4. Performance optimization. It is necessary to investigate methods of model optimization and data processing to improve performance and resource efficiency.
5. System deployment. An important point will be to explore ways to deploy a voice recognition system in a real environment and integrate it with other systems or platforms.
6. Research of new algorithms. To improve the model, new methods and algorithms used for voice recognition can be explored, such as deep neural networks or machine learning methods.

Overall, the voice gender recognition project has the potential for further development and improvement. Using the Python programming language and the Visual Studio Code integrated development environment provides flexibility, productivity, and access to powerful libraries and tools that support speech recognition development and research.

In this project, we have developed a voice-based gender recognition system that can have a wide range of applications and benefit people in many ways. With the help of this system, users can transmit textual information by voice, which can then be processed and used for various purposes. One of the main advantages of this project is convenience. Instead of typing text manually, users are allowed to simply speak and their voice will be converted into text. This is especially useful for people who have limitations in writing or writing speed, and for situations where hands are busy or unavailable. Also, our system can be useful in the field of accessibility. It can help people with visual impairments or other physical disabilities receive information by simply speaking. This makes information more accessible and helps improve their lives. In addition, our project can find applications in automated systems that require processing of a large amount of text. For example, government agencies, journalists, researchers or companies that work with a lot of documents can use this system to automatically recognize information from audio recordings. In general, our project helps to make the processing of text information more convenient, fast and accessible. It opens new opportunities for people and helps to automate routine tasks. We believe that our project can positively affect people's lives and make a significant contribution to the field of technology and information access.

Various algorithms and methods were used in our voice gender recognition project to achieve the goal. Starting with processing audio data using the *PyAudio* library, we recorded audio signals and applied normalization, cut out silence and added pauses. Next, we used the *librosa* library to extract such features from audio recordings as MFCC, Chroma, MEL Spectrogram Frequency, and others.

These features were used as input to train the gender recognition model. We used a neural network model using the *TensorFlow* library and trained the model with the help of a dataset divided into testing, validation, and training datasets.

As a result of our project, we have achieved the task of voice gender recognition with satisfactory accuracy. Our methods and algorithms allow for identifying signs of gender in human voices and carrying out classification. This project can be very useful in many fields, such as voice command recognition, virtual assistants, fraud detection, and other areas where gender recognition by voice can find its application.

The perspectives of further research should be an analysis of the advantages and disadvantages of gender classification models when recognizing an article by voice, taking into account accents, different qualities of the voice, indistinctness of speech and the imposition of background noise, the influence of emotion on the voice (anger, scream, joy, laughter, etc.), singing, speech speaker's text, nationality (some representatives of the nation speak too high or too low for the native speaker in other non-native languages). It will also be interesting to conduct a study of gender classification models in article-by-voice recognition on different test data sets and test these results in several different environments, for example, in schools, universities, marketing companies, and security systems. The only problem is getting permission from the participants to voluntarily consent to record their voices, so the problem of continuing the research is directly with confidentiality, consent and bias. This is especially important when dealing with sensitive data such as voice recordings.

Acknowledgement

The research was carried out with the grant support of the National Research Fund of Ukraine "Information system development for automatic detection of misinformation sources and inauthentic behaviour of chat users ", project registration number 187/0012 from 1/08/2024 (2023.04/0012). Also, we would like to thank the reviewers for their precise and concise recommendations that improved the presentation of the results obtained.

References

[1] Becker K (2017) Dataset, https://www.kaggle.com/datasets/primaryobjects/voicegender?resource=download
[2] Becker K (2016) Identifying the Gender of a Voice using Machine Learning, https://www.primaryobjects.com/2016/06/22/identifying-the-gender-of-a-voice-using-machine-learning/
[3] Bernstein JG, Stakhovskaya OA, Jensen KK, Goupell M J (2020) Acoustic hearing can interfere with single-sided deafness cochlear-implant speech perception. Ear and hearing 41(4):747–761. https://doi.org/10.1097/AUD.0000000000000805
[4] Bisikalo O, Boivan O, Khairova N., Kovtun O, Kovtun V (2021) Precision automated phonetic analysis of speech signals for information technology of text-dependent authentication of a person by voice. CEUR Workshop Proceedings 2853:276–288, https://ceur-ws.org/Vol-2853/paper34.pdf
[5] Brown LM (2022) Gendered artificial intelligence in libraries: Opportunities to deconstruct sexism and gender binarism. Journal of Library Administration 62(1):19–30. https://doi.org/10.1080/01930826.2021.2006979
[6] Brownlee J (2020) Support Vector Machine, https://machinelearningmastery.com/support-vector-machines-for-machine-learning/
[7] Cao YT, Daumé III H (2021) Toward gender-inclusive coreference resolution: An analysis of gender and bias throughout the machine learning lifecycle. Computational Linguistics 47(3):615–661. https://doi.org/v10.1162/coli_a_00413
[8] Chen X, Li Z, Setlur S, Xu W (2022) Exploring racial and gender disparities in voice biometrics. Scientific Reports 12(1):3723. https://doi.org/10.1038/s41598-022-06673-y
[9] Chouchane O, Panariello M, Zari O, Kerenciler I, Chihaoui I, Todisco M, Önen M (2023) Differentially private adversarial auto-encoder to protect gender in voice biometrics. ACM Workshop on Information Hiding and Multimedia Security, 127–132. https://doi.org/10.1145/3577163.3595102
[10] Cornacchia M, Papa F, Sapio B (2020) User acceptance of voice biometrics in managing the physical access to a secure area of an international airport. Technology Analysis & Strategic Management 32(10):1236–1250. https://doi.org/10.1080/09537325.2020.1758655
[11] DataSet (2023) https://raw.githubusercontent.com/primaryobjects/voice-gender/master/voice.csv
[12] Dumpala SH, Dikaios K, Rodriguez S, Langley R, Rempel S, Uher R, Oore S (2023) Manifestation of depression in speech overlaps with characteristics used to represent and recognize speaker identity. Scientific Reports 13(1):11155. https://doi.org/10.1038/s41598-023-35184-7
[13] Extracting features/ labels from .wav file to .csv (2023) https://stackoverflow.com/questions/53301682/extracting-features-labels-from-wav-file-to-csv
[14] Fenu G, Marras M (2022) Demographic fairness in multimodal biometrics: A comparative analysis on audio-visual speaker recognition systems. Procedia Computer Science 198:249–254. https://doi.org/10.1016/j.procs.2021.12.236
[15] Fenu G, Marras M, Medda G, Meloni G. (2021) Fair voice biometrics: Impact of demographic imbalance on group fairness in speaker recognition. Interspeech, International Speech Communication Association, 1892–1896. https://doi.org/10.21437/Interspeech.2021-1857
[16] Franzoni V (2023) Gender Differences and Bias in Artificial Intelligence. Gender in AI and Robotics: Intelligent Systems Reference Library 235: 27–43. https://doi.org/10.1007/978-3-031-21606-0_2
[17] Garain A (2020) Gender Recognition from Voice, https://ieee-dataport.org/documents/gender-recognition-voice
[18] Gender Recognition by Voice | 01 | Data Exploration (2023) https://matthew-nm.github.io/pages/projects/gender01_content.html
[19] Gender Recognition by Voice (2020) https://colab.research.google.com/github/shestakoff/hse_se_ml/blob/master/2020/s07-dimred/seminar7-homework.ipynb
[20] Gevin R (2022) Gender Recognition by Voice, https://rpubs.com/ReynaldiGev15/gender-recognition

[21] Goodfellow I, Bengio Y, Courville A (2016) Deep Learning. MIT Press, Cambridge, MA, USA
[22] Hamid HRMH, Nordin NW, Abdullah NY, Ismail WHW, Abdullah D (2024). Two Factor Authentication: Voice Biometric and Token-Based Authentication. Applied Problems Solved by Information Technology and Software 27–35. https://doi.org/10.1007/978-3-031-47727-0_4
[23] Hanifa RM, Isa K, Mohamad S (2021) A review on speaker recognition: Technology and challenges. Computers & Electrical Engineering 90:107005. https://doi.org/10.1016/j.compeleceng.2021.107005
[24] Hipólito I, Winkle K, Lie M (2023) Enactive artificial intelligence: subverting gender norms in human-robot interaction. Frontiers in Neurorobotics 17:1149303. https://doi.org/10.3389/fnbot.2023.1149303
[25] IBM. Random Forest (2023) https://www.ibm.com/topics/random-forest
[26] IBM. What is a Decision Tree? (2023) https://www.ibm.com/topics/decision-trees
[27] Iloanusi ON, Mbah CC, Ejiogu U, Ezichi SI, Koburu J, Ezika IJ (2019) Gender and age group classification from multiple soft biometrics traits. International Journal of Biometrics 11(4):409–424. https://doi.org/10.1504/IJBM.2019.102883
[28] Meena T, Sarawadekar K (2020) Gender recognition using in-built inertial sensors of smartphone. REGION 10 CONFERENCE (TENCON), 462-467. https://doi.org/10.1109/TENCON50793.2020.9293797
[29] Jurafsky D, Martin JH (2020) Speech and Language Processing (3rd ed.), https://web.stanford.edu/~jurafsky/slp3/
[30] Kao CY, Chueh HE (2023) Voice Response Questionnaire System for Speaker Recognition Using Biometric Authentication Interface. Intelligent Automation & Soft Computing 35(1):913–924. https://doi.org/10.32604/iasc.2023.024734
[31] Kholodna N, Vysotska V, Albota S (2021) A Machine Learning Model for Automatic Emotion Detection from Speech. CEUR Workshop Proceedings 2917:699–713, https://ceur-ws.org/Vol-2917/paper42.pdf
[32] Kobylyukh L, Rybchak Z, Basystiuk O (2023) Analyzing the Accuracy of Speech-to-Text APIs in Transcribing the Ukrainian Language, CEUR Workshop Proceedings 3396:217–227, https://ceur-ws.org/Vol-3396/paper18.pdf
[33] Luna JC (2022) Choosing Python or R for Data Analysis? An Infographic, https://www.datacamp.com/community/tutorials/r-or-python-for-data-analysis
[34] Mewada A (2018) Gender Recognition by Voice Acoustic Parameters, https://www.kaggle.com/code/akshaymewada7/gender-recognition-by-voice-acoustic-parameters
[35] Nunes IRDAL (2021) A Conceptual Model for Gender-Inclusive Requirements (Doctoral dissertation), https://run.unl.pt/bitstream/10362/145198/1/Nunes_2021.pdf
[36] Perron M, Dimitrijevic A, Alain C (2022) Objective and subjective hearing difficulties are associated with lower inhibitory control. Ear and Hearing 43(6):1904–1916. https://doi.org/10.1097/AUD.0000000000001227
[37] Prodi N, Visentin C, Borella E, Mammarella IC, Di Domenico A (2019) Noise, age, and gender effects on speech intelligibility and sentence comprehension for 11-to 13-year-old children in real classrooms. Frontiers in psychology 10:2166. https://doi.org/10.3389/fpsyg.2019.02166
[38] Rabiner L, Juang B (2006) Speech Recognition, Statistical Methods. https://web.ece.ucsb.edu/Faculty/Rabiner/ece259/Reprints/355_statistical%20speech%20recognition%20proof.pdf
[39] Rincón C, Keyes O, Cath C (2021) Speaking from experience: Trans/non-binary requirements for voice-activated AI. Proceedings of the ACM on Human-Computer Interaction 5:1–27. https://doi.org/10.1145/3449206
[40] Rusyn B, Chorniy A (2008) Application of wawelet-transformation in to the system of speech recognition. Modern Problems of Radio Engineering, Telecommunications and Computer Science, 345. https://ieeexplore.ieee.org/abstract/document/5423385
[41] Sazhok M, Poltieva A, Robeiko V, Seliukh R, Fedoryn D (2021) Punctuation Restoration for Ukrainian Broadcast Speech Recognition System based on Bidirectional Recurrent Neural Network and Word Embeddings. CEUR Workshop Proceedings 2870:300–310, https://ceur-ws.org/Vol-2870/paper25.pdf
[42] Shakhovska N, Basystiuk O, Shakhovska K (2019) Development of the Speech-to-Text Chatbot Interface Based on Google API. CEUR Workshop Proceedings 2386:212-221, https://ceur-ws.org/Vol-2386/paper16.pdf
[43] Markowitz JA (2000) Voice biometrics. Communications of the ACM 43.9:66-73. https://dl.acm.org/doi/pdf/10.1145/348941.348995
[44] Si S, Li Z, Xu W (2021) Exploring demographic effects on speaker verification. In 2021 IEEE Conference on Communications and Network Security (CNS) (pp. 1–2. https://doi.org/10.1109/CNS53000.2021.9729038
[45] spafe.features.spfeats (2019) https://spafe.readthedocs.io/en/latest/features/spfeats.html
[46] SpeechRecognition (2022) SpeechRecognition 3.10.1, https://pypi.org/project/SpeechRecognition/
[47] Srivastava T (2018) K-Nearest Neighbors, https://www.analyticsvidhya.com/blog/2018/03/introduction-k-neighbours-algorithm-clustering/
[48] Sruthi ER (2021) Random Forest, https://www.analyticsvidhya.com/blog/2021/06/understanding-random-forest/
[49] Swaminathan S (2018) Logistic Regression, https://towardsdatascience.com/logistic-regressiondetailed-overview-46c4da4303bc
[50] Szlavi A, Guedes LS (2023) Gender Inclusive Design in Technology: Case Studies and Guidelines. Lecture Notes in Computer Science 14030:343–354. https://doi.org/10.1007/978-3-031-35699-5_25
[51] Tarres Puertas MI, Merino Millo J, Dorado Castaño AD (2021) Qu-Bot-H2O challenge: Integration of computational thinking with chemical experimentation and robotics through a web-based platform for early ages including gender, inclusive and diversity patterns. International Conference of Education, Research and Innovation, 8361–8365. https://doi.org/10.21125/iceri.2021
[52] Tejale SS, Kute TB (2020) Performance Evaluation of Algorithms for Gender Classification, https://www.ijeast.com/papers/568-573,Tesma411,IJEAST.pdf
[53] Trysnyuk V, Nagornyi Y, Smetanin K, Humeniuk I, Uvarova T (2020) A method for user authenticating to critical infrastructure objects based on voice message identification. Advanced Information Systems 4(3):11–16. https://doi.org/10.20998/2522-9052.2020.3.02
[54] Tymoshenko K, Vysotska V, Kovtun O, Holoshchuk R, Holoshchuk S (2021) Real-time Ukrainian text recognition and voicing. CEUR Workshop Proceedings 2870:357–387, https://ceur-ws.org/Vol-2870/paper27.pdf
[55] Vaidya J, Gujar S, Devani H., Makhija R, Naik D (2017) Voice Recognition, https://prezi.com/36iofn4i71oy/voice-recognition/
[56] Verma R (2018) Detailed parameter analysis and tuning, https://www.kaggle.com/code/deadskull7/detailed-parameter-analysis-and-tuning-97-004

[57] Voice Gender (2021), https://github.com/primaryobjects/voice-gender/blob/master/readme.md
[58] Wu E, Li Z, Xu W (2021) Voice Doppelgänger Susceptibility among Racial and Gender Groups: IEEE CNS 21 Poster. Communications and Network Security, 1–2. https://doi.org/10.1109/CNS53000.2021.9729035
[59] Yalova K, Babenko M, Yashyna K (2023) Automatic Speech Recognition System with Dynamic Time Warping and Mel-Frequency Cepstral Coefficients. CEUR Workshop Proceedings 3396:141-151, https://ceur-ws.org/Vol-3396/paper11.pdf
[60] Zäske, R, Skuk VG, Golle J, Schweinberger S R (2020). The Jena Speaker Set (JESS) –A database of voice stimuli from unfamiliar young and old adult speakers. Behavior research methods 52:990–1007. https://doi.org/10.3758/s13428-019-01296-0
[61] Abbas, S., Alsubai, S., Sampedro, G.A., Abisado M., Almadhor A. S., Kryvinska, N., Zaidi, M.M. (2023) Active Learning for News Article's Authorship Identification. IEEE Access 11, 98415–98426. https://doi.org/10.1109/ACCESS.2023.3310813
[62] Abbasi, A., Javed, A.R., Iqbal, F., Jalil Z., Gadekallu, T.R., Kryvinska, N. (2022) Authorship identification using ensemble learning. Scientific Reports 12(1), 9537. https://doi.org/10.1038/s41598-022-13690-4
[63] Teslyuk V, Tsmots I, Kryvinska N, Teslyuk T, Opotyak Y, Seneta M, Sydorenko R. (2023). Neuro-controller implementation for the embedded control system for mini-greenhouse. PeerJ Computer Science 9:e1680 https://doi.org/10.7717/peerj-cs.1680
[64] Vysotska, V., Chyrun, L., Chyrun, S., & Soltys, M. (2024). Information technology for textual content author's gender and age determination based on machine learning. CEUR Workshop Proceedings 3723: 498-540, https://ceur-ws.org/Vol-3723/paper27.pdf.

Authors' Profiles

Victoria Vysotska is an Associate Professor at the Information Systems and Networks Department of Lviv Polytechnic National University. She defended her Doctoral degree in Technical Science, speciality in «structural, applied and mathematical linguistics» in 2023 on the topic "Analysis and synthesis of computational linguistic systems for processing Ukrainian textual content" Also, she received a PhD degree in Information Technologies from Lviv Polytechnic National University in 2014. She has currently published more than 50 publications. Her main research interests are focused on the identification of disinformation/fakes/propaganda, detection of sources of disinformation and inauthentic behaviour (bots) of coordinated groups based on artificial intelligence, NLP, computer linguistics, data science, system analysis, information technologies, machine learning, cyber security, modern education and distance learning system development.

Denys Shavaiev is a dedicated student enrolled in the Department of Information Systems and Networks at Lviv Polytechnic National University. His academic journey is fuelled by a profound interest in cutting-edge technologies such as Machine Learning, Artificial Intelligence, and Data Science.

Michal Greguš is a professor at the Faculty of Management, Comenius University Bratislava. He finished his university studies with summa cum laude and obtained his PhD degree in the field of mathematical analysis at the Faculty of Mathematics and Physics, Comenius University in Bratislava. He has been working previously in the field of functional analysis and its applications. At present his research interests are in applied mathematics, modelling of economic processes and data and business analytics.

Yuriy Ushenko is a professor in the Computer Science Department at Chernivtsi National University, Chernivtsi, Ukraine. Research Interests: Data Mining and Analysis, Computer Vision and Pattern Recognition, Optics & Photonics, Biophysics.

Zhengbing Hu: Prof., Deputy Director, International Center of Informatics and Computer Science, Faculty of Applied Mathematics, National Technical University of Ukraine “Kyiv Polytechnic Institute”, Ukraine. Adjunct Professor, School of Computer Science, Hubei University of Technology, China. Visiting Prof., DSc Candidate in National Aviation University (Ukraine) from 2019. Major research interests: Computer Science and Technology Applications, Artificial Intelligence, Network Security, Communications, Data Processing, Cloud Computing, Education Technology.

Dmytro Uhryn graduated from Yuriy Fedkovych Chernivtsi National University, Chernivtsi. Currently, he is a Doctor of Technical Sciences and associate professor at Yuriy Fedkovych Chernivtsi National University. He has currently published more than 170 publications. His research interests are data mining, information technologies for decision support, swarm intelligence systems, and industry-specific geographic information systems.

How to cite this paper: Victoria Vysotska, Denys Shavaiev, Michal Greguš, Yuriy Ushenko, Zhengbing Hu, Dmytro Uhryn, "Information Technology for Gender Voice Recognition Based on Machine Learning Methods", International Journal of Modern Education and Computer Science(IJMECS), Vol.16, No.5, pp. 65-87, 2024. DOI:10.5815/ijmecs.2024.05.05