**Using machine learning algorithms to classify the Hamzat Alwasl and Hamzat Alqatae**

**Abstract**

The Arabic language is considered a living organism that grows and develops through the correct practice and application of all its literature and its syntactic, morphological, semantic, and lexical branches. In this research study, we review the contribution of technology to the development of Arabic language, especially the correct writing of Hamzat Alwasl and Hamzat Alqatae. This project aims to build a smart Classifier model to classify Arabic words beginning with the letter alif and their Hamzat Alwasl and Hamzat Alqatae through the using of artificial intelligence techniques in general and machine learning algorithms to establish accurate and correct criteria in writing the Hamzat Alwasl and Hamzat Alqatae correctly.

Consequently, technology would be adapted to contribute to the service of Arabic language.

This research relied on the compilation of Arabic words beginning with hamza by designing a digital questionnaire. The task of this questionnaire is to collect the largest number of words beginning with Hamza and to classify them as Hamzat Alwasl and Hamzat Alqatae according to the grammatical rules followed in this process. The questionnaire was circulated in the internet web and was filled out by fifty specialists in syntax with different academic ranks. The total number of classified words reached 400 words, and after processing and excluding the repeated words, 101 words, we obtained 299 valid words to be applied to the Classifier model, and based on the size and type of the collected data and the classification mechanism followed, classification algorithms were applied that fit the collected sample such as the vector support machine algorithm , Naif Biz algorithm, and Nearest neighbor algorithm by using Python language and sk-learn library.

After training the used classificatory models and measuring the accuracy of the algorithms, it was quite apparent that the vector support machine (SVM) algorithm had obtained the highest accuracy of 92%, which is high and sufficient to solve the problem of the project.

**preamble:**

We can define machine learning as following the science of artificial intelligence that is concerned with designing algorithms and techniques that allow computers to process self-learning with the possibility of developing these algorithms.

We can review two types of machine learning: inductive learning and deductive learning, inductive learning is known as inductive learning and works on the principle of also deducing general rules from data, and this differs from inductive learning where the learner is given the rules that he needs to apply.

Machine learning depends on the principle of systems learning from available data, identifying appropriate patterns and making decisions without human intervention, and since the primary task of machine learning is to extract information from data, machine learning is a method of data analysis that automates building data analysis models known as analysis models. Predictive, this pattern is known as the predictive or taxonomic analysis model, these models allow researchers, data scientists, and analysts to make decisions that enable reliable results.

Building a data analysis model is subject to several mechanisms or ways to reach accurate results, as the model building stage requires the implementation of several arranged and sequential mathematical or logical steps to reach a trained analytical model that can predict values ​​closer to the truth or accurately classify the data. This group of steps is called the name Algorithm, a data classification model is also called a classifier. In this project, I will address the use of machine learning algorithms models in classifying the Hamzat Alwasl and Hamzat Alqatae, and these algorithms will be applied to Arabic words because the Arabic language has not been as well studied and researched as other languages.

**Project objectives:**

* Solving the problem of the wrong use of the hamzat "Wasl" and "Gtaa" in Arabic texts. Identify the most appropriate classification algorithms within machine learning algorithms to determine the correct positions, using the "Hamzat Wasl" and "Hamzat Gtaa" at the beginning of a word.
* Measuring the quality of the most famous classification algorithms in distinguishing between the "Hamzat Wasl" and "Hamzat Gtaa "at the beginning of a word.

**The problem of the project and itsquestions:**

The main problem of the project lies in the erroneous use of misplaced link or cutting hums when writing words with the letter A in particular, where many mistake the drawing or dropping of the whisper, which weakens the power of thelanguage. This project contributes to the development of the Arabic language, by building a smart classification model for editors and practitioners, and all language users from the correct classification of the link and cutting. This research seeks to answer the following questions in light of the use of machine learning algorithms used in classification and prediction:-

* How can we distinguish between the two links and pieces in Arabic texts?
* What are the benefits of classifying the links and pieces in Arabic texts?
* What machine learning algorithms are suitable for classifying the link and cutting hums in Arabic texts?

**Previousstudies:**

This study is one of the new studies in the field of grammatical and technical studies dealing with the use of machine learning algorithms to classify the Hamzat Alwasl and Hamzat Alqatae in Arabic texts, and as far as I know, and through my research, I have not found any studies that have been exposed to this subject before because of the difficulty of using software libraries in dealing with Arabic texts deservedly, but there is a great similarity and overlap of other studies associated with the project to use these algorithms in similar and interesting applications, and I have benefited from them, and helped me enter my project, including:

**Study entitled:**

* Exploration of opinions in arabic comparison sentences **[1]**

This study addressed the problem of identifying the comparison magazine in the exploration of opinions used in the Arabic text, and the study focused on drawing opinions from the comparison magazine by knowing the product preferred by the opinion writer compared to another product or more. The researcher stated that there is some research in this field for the sentences of English and its languages, but for Arabic sentences this is the first study, and the study also used a technique based on linguistic classification and other technique based on machine education.

**Study entitled:**

* Comparative study of opinion exploration algorithms and analysis of emotions and theirapplications **[2]**

This study addressed the problem of the multiplicity of views of customers stored in the data repositories of the Internet, which gave data exploration and sentiment analysis attention in recent years, the researcher stated that people have relied on the machine in classifying and processing data, as the availability of huge amounts of views on a single product helps to predict customer feelings by analyzing views that help not only increase profits but also improve The product this study compared the technologies currently available and used in multiple applications in the field of exploration ofopinions. After reviewing the studies above, the idea of classifying the humza was based on the use of the concept of exploration in Arabic texts by researchers in the above studies using the method of prospecting in Arabic texts, with different classification algorithms used.

**Project methodology:**

* In this project, digital resolution was used, and the resolution task was limited to compiling and classifying as many start-up words as possible by Arabic language specialists (for two links or pieces in accordance with the grammatical rules of the classification, the link and the pieces.
* The statistical approach based on the designed resolution and the word training group was also used in this project using the classification model, and for this project we will use machine learning algorithms after considering the frequency of conditions and characteristics (grammars associated with the words started with the whisper and extracted via resolution) to deduce the basic elements of the classificationmodel.

**Project structure:**

**The project consists of threeinvestigations:**

* The first research (theoretical framework of the project): deals with the concept of artificial intelligence, the concept of machine learning, data classification algorithms and the link and cuttingwhispers.
* The second research (accelerator and application): deals with the stages of application of works models, which include: collection of descriptions, cleaning, rectifiers, coding of data, identification of independent and dependent variables in the data sample, as well as the construction of worksmodels.
* The third research (results of the project outputs): The test of works models, measurement of model accuracy, results, project outputs and the benefits of theproject.

**First topic: The theoretical framework of the project:**

**First: The concept of artificialintelligence:**

Artificial Intelligence is a branch of computer science, one of the pillars of the technology industry of our modern era, referredto asAI and can be defined as the ability of machines and computers to perform tasks that largely mimic the intelligent human mind, and these tasks can be summarized in the ability to think or learn from its previous experiences, so we can say that artificial intelligence aims to reach systems that behave [3] They learn and understand as they behave, learn and understand human beings as they possess**intelligence.[3]**

**Types of artificialintelligence:**

Ai types can be divided according to their capabilities into three **types[4]** as follows:

* **Limited artificialintelligence:**

It is a type that can perform specific and clear tasks such as self-driving car applications, speech recognition software, images or chess, and this type of artificial intelligence is the most common type.

* **General ArtificialIntelligence:**

It is a species that has thinking capabilities similar to that of man, as it makes the machine be able to think on its own and is very similar to human thinking, and in fact there are no practical applications for this species, but there are only research studies that need a lot of effort to turn it into reality, and the method of neural networks is one of the models of general artificial intelligence, as it is concerned with the production of a system of neural networks of the machine similar to that contained in the human mind. **[5]**

* **Unlimited artificialintelligence:**

Unlimited artificial intelligence is the kind that may exceed the level of human intelligence, and its ability to perform tasks in a way that can be better than that of specialized and knowledgeable humans, and this type has many necessary characteristics, such as learning, automatic planning, the ability to communicate and make the right decision, but the concept of super-ARTIFICIAL INTELLIGENCE is a hypothetical concept that does not exist in our time.

**Artificial intelligence can also be classified according to the following four differentfunctions:**

* **Interactivemachines:**

It is the simplest type of ARTIFICIAL INTELLIGENCE because it lacks the ability to learn from previous experience to develop future business, so here we will interact with the current experience to produce the best possible way, such as blue deep equipment developed by IBM and AlphaGo from Google.

* **Limited memory:**

Limited Memory can store previous historical data on the current system for a limited time period, and the self-driving approach is one of the most effective examples of this pattern, maintaining the last speed of other cars, the distance between those cars, the speed limit, and other basic information for driving through transport routes. **[6]**

* **Theory of reason:**

This type of artificial intelligence means that the machine absorbs human feelings, interacts with and communicates with humans, and it should be noted that no practical applications have been found to date in that type of artificialintelligence.

* **Self-perception:**

Awareness-Self is a future prediction for artificial intelligence, and works on a very modern technical and sensory principle where the machine can generateself-knowledge and sensations of its own, which will make it smarter than the human being, and that concept is still not present in fact. **[7]**

**Artificial intelligence subfields:**

AI has many sub-areas, such as machine learning, which includes enabling computers to learn independently of any previous experience, so that computers can predict the right decision-making quickly, through the development of an algorithm that allows for this situation. It should be noted that this term was first proposed by Arthur Samuel in 1959 and we will refer below to some of the most famous areas of artificial intelligence as follows:

* **Data exploration:**

It is intended to search and explore specific data and patterns within a wide range of data through software, as companies can benefit from data exploration in developing their performance, increasing sales and reducing production costs. **[8]**

* **Information retrieval and semantic web:**

The concept of retrieving information refers to the process of searching for any type of data and documents that may be present on the Internet through the semantic web concept. The semantic web service converts web data into a global database of interconnected information, so that machines can understand it, and not only use it for humans. In this way, the machine can book tickets online, use an online dictionary, or other things that initially require manual use to complete them. **[9]**

* **Representation ofknowledge:**

Knowledge representation is an area of artificial intelligence that is interested in making machines think and make the right decision, as the knowledge acquired by the machine is grouped and stored in a database that is used to share knowledge and manage its components, and is a reference for making any smart decisions made by the machine in thefuture.

* **Logical thinking and probability thinking:**

Logical thinking in artificial intelligence is a different form of thinking, because facts are inferred based on available data. Logical thinking corresponds to the so-called probability thinking, which uses concepts of probability and uncertainty in knowledge to deal with all future uncertainties of all events that may be suspected. **[10]**

* **Machine learning:**

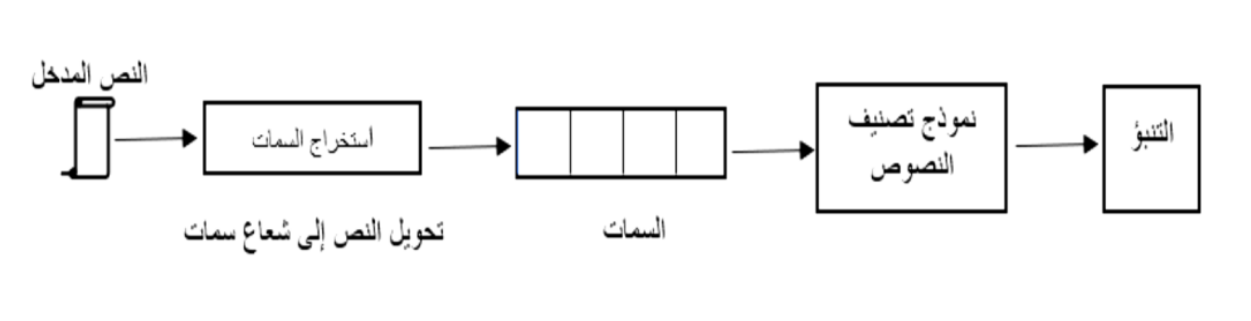
Machine learning is a branch of artificial intelligence, which includes the design and development of algorithms and techniques that allow computers to possess "learning" properties. In general, learning is divided into two levels: inductive and introspective, where the introspective approach derives general rules and provisions from bigdata.

**Second: Machine Learning:**

The main task of machine learning is to extract valuable information from data, so it is very close to data extraction, using machine learning in data analysis, a way to develop complex models and appropriate algorithms to extract data using predictive processes, called predictive analysis. These analytical models allow researchers and data analysts to learn reliable decisions and results and can recognize stored data and theirrelationships.

Machine learning systems can also be defined as systems that make predictions based on what previous data have learned and need to be trained in many examples of text and forecasts (markers) expected for each of them. The data used for training are called the training data set and these data are pre-classified with features and the more accurate the training group and the selected features are appropriate, the better theclassification predictions, when a worked-by-machine learning method is trained, training data must be converted into something you can understand. The machine, where features are extracted and converted into rays (representation of text by numbers) will help it learn from existing data and make predictions about future texts. **[11]**

The trained model can extract features from new text, predict or categorize texts by specific characteristics using data classification algorithms as in figure (1-1) below:



Form (1-1): Demonstrates the text classification mechanism using data classification algorithms.

**Third: Data classification algorithms:**

There are several data classification algorithms that fit applications for text data exploration with ease after being reviewed, and Anya is easy to train with both large and small amounts of data provided, and here we will review the most famous machine learning algorithms for classifying text data that have been used in this project:

1. **خوارزمية دعم المتجه (SVM-Support Vector Machines):**

Known asSVM, this algorithm is based on machine-oriented learning algorithms based on a pre-defined data set (training group) in the algorithm's training so that it can analyze, classify or identify any new set of data, developed by worlds Vladimir Fabank and Alex Shrivonenkis in 1963 and then developed by Corina Kurtz and Fabank in 1993 and published in 1995. **[12]**

The SVM algorithm is one of the most popular automated classification methods based on the creation of a curve or break level, separating samples entered from each other, and the algorithm is characterized by its use in classifying issues with binary categories exclusively. For classification and pattern identification, one of its features is high accuracy in classification and is applied in a wide range of areas, including the selection of text categories by image classification. **[13]** If the algorithm can find a break level one lower than the point vectors, the classification is linear, otherwise the classification is non-linear. **[14]**

The accuracy of the algorithm determines its ability to separate two types so that the two closest points to each other are far from possible, and we can name the dividing level with the edge or margin of separation. In general, the higher the edge or margin of separation, the less error when generalizing to a new dataset.

1. **خوارزمية نايف بيز: (Bayes Naive)**

The Nayef Biz (NB) algorithm is also a machine-oriented learning algorithm, and relies on the rules of probability formulated by scientist ThomasBiz[15] where the probability is calculated using the number of repetitions of values, frequencies and value combinations in pre-known data (yip data). It is consideredthe baseline in the classification of texts and is effective in many areas although there are a number of other works with higher accuracy such as theSVM model, as the Bayes Naive model distributes texts for each item using a probability model with independent assumptions, this method is very common in the field of text classification, as the binary work is one of the best known methods of the Bayes Naive model that used two-value radial representation of texts.

Several improvements have been made to the Biz class, including adjusting the probability calculation, reducing attributes and a few other characteristics, and as Biz's theory looks for the possibility of an event, another event has already occurred. Nayef Baze's algorithm is also one of the most important machine learning algorithms for several reasons: the ease of building a work, and here you do not need to use so-called schemes estimation for any complex repetitive variables, and may be easily applied to a large data set and the goal of the algorithm is to build a base that allows future structures to be allocated to a particular item by giving vectors to the variables describing that structure, by which the user can make many statistics for ease of use. **[16]**

1. **خوارزمية الجار الأقرب: K-Nearest Neighbor**

KnN can be used as a simple and effective text classification work, knn contains two disadvantages: computational complexity if the samples are very similar, and their performance is easily affected if the training samplesareindividual. KnN's complexity can be reduced by using three methods: either by reducing the dimensions of the vector representing the text, by reducing the amount of training samples or by limiting the finding of the nearest neighbors, i.e. the value of the K.

KNN td uses text classification by calculating the distance the text builds and all texts in the training data set using a measure of difference or similarity between them, then finding the nearest K adjacent to all training texts and selecting the text class to the item with the largest number of texts in the nearest neighbors of texts and other algorithms has been improved in more ways than one. **[17]**

**Fourth: They whispered the connection and the pieces**

In this part of the theoretical aspect, we will address the link and cutting whispers in the first word under the followingnames:

* **Meaning of humming (language):**

Many dictionaries of the language, including the tongue of the Arabs of Ibn Mansaer, stated that the humming in language means winking, winking and pressing. **[18]**

* **The meaning of humming (in terms of):**

Al-Hamza in the term linguists received several definitions, including the definition of Al-Azhari in his book The Refinement of language, where he says " I know that the whisper does not satire, but writes once thousand and once Wawa and once J and thousand soft does not have a letter, but it is part of the period after opening, and the letters twenty-eight letters with W and Alpha and Z, and are carried out with a whisper twenty-nine letters and a whisper as the correct letter, However, it has cases of softening, deletion, substitution and mitigation... It is not a hollow letter, it is a ring from the far mouth."

**The sections of the humming at the beginning of theword:**

At the beginning of the word, the humming is divided into mysection:

**A- Thelink:**

* **Nameit:**

The link is named after this name, because it reaches the pronunciation of the static letter located at the beginning of the speech, since the important rule in that is that it does not start with a resident and does not stop with amovement.

* **Defined:**

It is the whisper that is fixed in the beginning, and falls in the case of the connection, and ibn Malik stated: "If it is called what is first a link, the whisper will be cut off if it is in a transfer from an act such as: (I saw I know)" is specified if a person is named after **him[19]**.

* **The movement of thelink:**

The original in the link is that, it is broken like: go, except for a thousand definitions that are open as well as a thousand right-handmen like you say: "Ayman Allah by opening thethousand." **[20]**

* **Hubpositions:**

The link has three positions asfollows:

**Names:**

In the first word, Hamza al-Wasl is located in ten names in the language (name, son, daughter, two, two, woman, son, im allah, est) and a number of these names are mentioned in the Qur'an.

**B-Actions:**

The link in the acts that are started is twoacts:

* The past act: the link in the past of the quintet is like: open up, as well as the past of the hexagons such as:Forgiveness.
* Do it: The link in doing so is in the followingplaces:
  + The order of the triple act is like: Victory- < victory.
  + The order of the five-year act is like: Triumph for thetruth.
  + The order of the six-party act is like: Forgive yourguilt.

**C-Letters:**

There is no link in the letters except in one letter, the induction (the) or what is known as solar pain, such as the sun, as well as in lunar lam such as: **moon[21]**

**B/ Thecutting** hum:

* **Nameit:**

The humza of the pieces is named after this name because it cuts and reserves the previous ones of the letters, such as your saying: Lesson - Study, the whisper cut the dal from the whisper that accepted it as well as Nasr - Anser, the whisper booked the next, which is the letter of the n for the whisper that acceptedit.

* **Cutting hummingpositions:**

The pieces have three positions asfollows:

**Names:**

The pieces are featured in all names in Arabic such as Ahmed, Amjad, Akram, Assaad,as well as thesix names, all that begins with a whisper are pieces such as: "Your brother-father", except for the 10 names whose whisper is a link.

**B-Actions:**

The pieces are placed in the verbs asfollows:

* + The past of the triangular act is like: take-eat.
  + The past of the quartet such as: Akram - Accomplished
  + The order of the four-way act is like:Akram.
  + The quadrant source is like: warning from analarm.

**C-Letters:**

All the letters in Arabic were shaken by a piece of humming, except for the introductory (N-N-N-SO-22 to)  **[22]**

* **The difference is my son, my whisper is the connection and the**pieces:

In order to distinguish and distinguish between the link and the pieces, it is necessary to know the following:

* The pieces come still ormoving.
* The link that comes onlymoving.
* The whisper of the pieces is fixed at the beginning of the speech and at the end of it, such as: he asked in the middle and read at the end.
* The link falls while speaking while only initially proven.
* The link is only excessive, while the pieces are original or **excess.[23]**

**Topic 2: Application and Implementation**

**Boot:**

Today's data swell day by day and their sources are numerous, and this leads to the exposure of these data to many problems that reduce the quality of data such as the large number of data lost and the inconsistency of data, so in this project I divided the stages of application of data works models into six stages, starting with the stage of resolution design and data collection and ending with the stage of measuring the accuracy of works models, as in the following figure (1.2):

****

**Form No. (1-2): Stages of application of works models**

In this research, we will address the first four stages of the application of works models and the remaining two phases will be addressed in the third.

* **The resolution design phase and data description.**

The arabic words that started with the whisper were compiled by designing a digital questionnaire see (annex A) and were designed using Google models and named (resolution of the classification of novice words with a link or cut), the task of resolution was limited to compiling as many start-up words as possible by Specialists in Arabic and were classified by specialists as linking or cutting in accordance with known grammar rules, and the questionnaire was published on the following web address:

[**https://docs.google.com/forms/d/e/1FAIpQLSdtwnflWQ7hoTdVbKfZUHDVix4fmXPE3grXMiQGIDem9Dsc0Q/viewform**](https://docs.google.com/forms/d/e/1FAIpQLSdtwnflWQ7hoTdVbKfZUHDVix4fmXPE3grXMiQGIDem9Dsc0Q/viewform)

It was filled by 50 grammar specialists with varying degrees with the aim of obtaining a homogeneous sample, and the total number of words classified was 400words. The words derived from the sample texts are categorized into words starting with a link referred to in this project by Wasl and words starting with a piece whisper referred to in this project.

Independent variables (characteristics) were divided into three characteristics in accordance with the grammar rules later referred to in table (1.2).

The dataset file format was selected as csv and contained characteristics (independent variables):

* Variable :(diacritic) is a digital variable with a length of (1) a number indicating the movement of the letter A at the beginning of the word (slot, fraction, bundle) represented by digital values (1,2,3) respectively.
* Variable :(Count) is a digital variable that refers to the number of word letters as the word in values containing two, three or four characters ... Etc., they are digitally represented in values (10... 4 or 3,2)respectively.
* The variable :(noun) is a digital variable that indicates that the word (not a name, a normal name, of the ten names, of the six names, a connected name, an active name) was represented by numbers (6, ... 3, 2, 1, 0, respectively.
* Variable :(verb) is a digital variable that indicates that the word (not effective, past act, act of action, action of order) has been digitally represented

(0,1, 2,3) respectively.

* The variable:(adjective) is a digital variable that indicates that the word (not adjective, characteristic) is digitally represented by numbers (0,1) respectively.
* The letter is a digital variable that indicates that theword (not a letter, a character) is represented in numbers digitally (0,1) respectively.
* The variable (the)is a digital variable that indicates that theword (not known as Al, known as Al) was represented by numbers digitally (0,1) respectively.
* The Dataset file is named Words Arabic,table (2-1) showing part of the profile data description (Words\_Arabic. csv) As follows:

**Picture containing a table

Description created automatically**

* **Data classification phase**

Certainly, when data quality is low, this will inevitably affect the results of the analysis. In this project, several methods of datacleaning wereused on the collected texts, and the data clean-up phase included the following stages:

* Dealing with lost data
* Delete duplicatedata.

After data cleaning on the collected texts, after accelerating and excluding duplicate words as well as completing missing data, the number of excluded words became 101, we got 299 words valid to apply to the workbook form.

Picture containing a table

Description created automatically

Table (2.2) explains the rules for locating the link and cutting hums.

The data in table (1-2) is encoded and literal values converted into digital values so that algorithms can handle them according toschedule 2-2, which explains the rules for locating the link and cutting whispers and the data after encoding it as in the table.

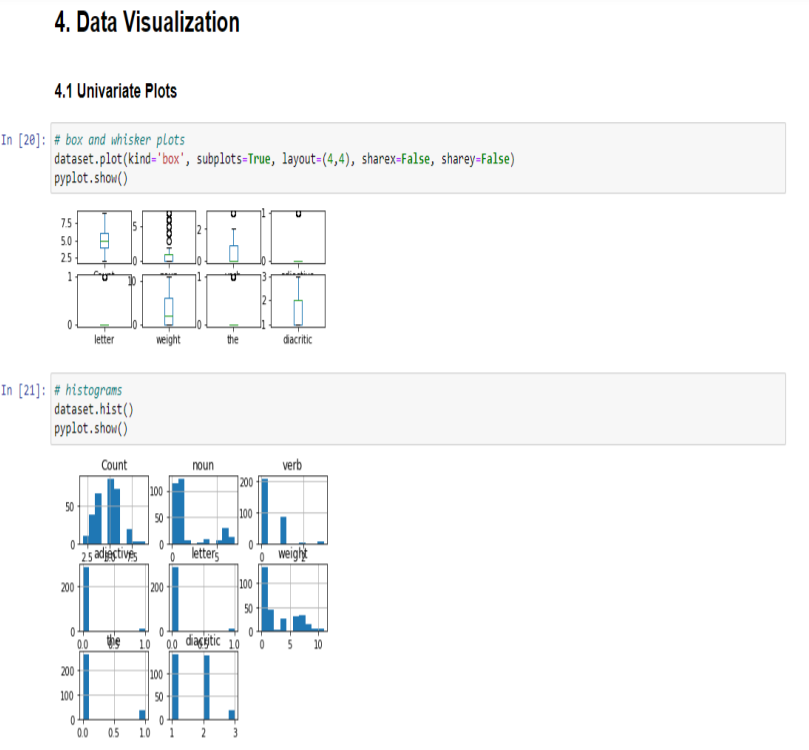
(3.2) The following:

Picture containing a table

Description created automatically

Table (2-3) Data coding in the data set file (Arabic.csv)

After the data coding process is completed, the data is represented using Python and the import of libraries (sklearn-scipy-numpy) using the editor (Jupyter) as in figure (2-2) as follows:



Shape (2-2 :)dyter data representation usingNotebook Jupyter)

* **Construction and training phase of worksmodels:**

The data represented were divided into training data (67%) and experimental data (33%) of the total data (299) recorded in preparation for the construction of the data training model using selected classification algorithms (SVM vector support algorithm - Naif Biz NB algorithm - nearest neighbor algorithm (KNN) selected relative to sample size and data values with the above algorithms and the model was built using theLearn-Sklibrary functions as in figure (2-3)a Below:

Picture containing text

Description created automatically

**Form (2-3): Build classification models using editor (Jupyter)**

**Third: Results and Outcomes of the Project**

**First: Testing works models**

After the completion of the construction phase of works models using algorithms ,(NB, SVM KNN) we moved to the test phase of these models so that we can measure their quality later, the models were tested by making predictions on the models after training using the library functions (sklearn),figure (3-1) shows the test of works models using the predict function as follows:

Picture containing text

Description created automatically

Form (3-1) shows the test of works models using predict function

**Second: Measuring the accuracy of works models**

The evaluation of the results of data exploration models is an important stage that enables us to define the most effective model, the effectiveness of the model is measured by the accuracy of the plan in place and the nature of the data used in the construction of models plays a key role in their effectiveness, and there are many statistical frameworks that test classification models, the most important of which are:

* **Calculating average accuracy average** accuracy

Is the mathematical average of the correct forecast accuracy ratios for each category provided by the model to the number of actual classifications of this category in the test data set, and the average accuracy of the predictions for the algorithms used as shown in table (3.1) which shows that the SVMalgorithm) has the highest ratio when calculating the average accuracy where it reached its accuracy (%94) while my algorithm (NB) and

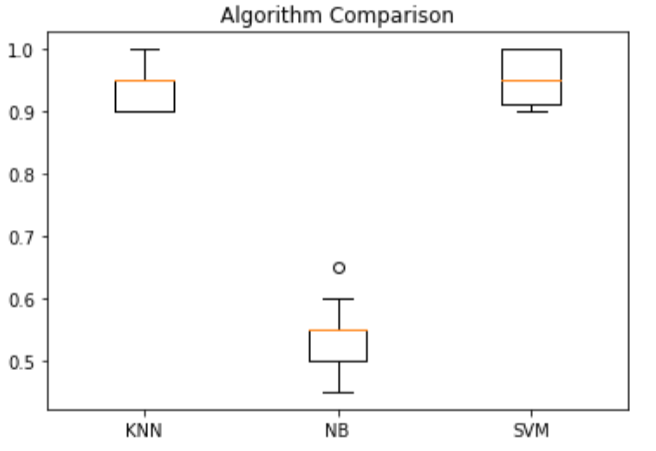
(KNN) On average accuracy (54%) and (85%) respectively:

Picture containing a table

Description created automatically

**Table (1-3): Average algorithm accuracy**

Figure (3-2) shows a graph of the average accuracy of algorithms (SVM), NB and KNN.



**Shape (2-3): Graph of average algorithm accuracy**

* **Accuracy Tota** Total Accuracy Calculation

The ratio of the total correct forecasts provided by the form to the total actual classifications in the test data set, the total accuracy of the algorithms used as shown in table 3.2)

Picture containing a table

Description created automatically

**Table (2-3) Total accuracy of algorithms**

From the table above, knn's algorithm has a total accuracy of 91% compared to the SVM algorithm, which had the highest overall accuracy rate of 92%, while the NB algorithm had the lowest overall accuracy rate of 78%.

* **Find matrix con** confusion matrix

The confusion matrix shows the number of cases correctly expected and the number of cases wrongly expected in the test group for each item compared to the number of actual cases of those items. The rank of this matrix is n × n where n is the number of items in the target variable column) the child variable, and the confusion matrix of the algorithms used to classify outcome variable has been found into two categories (wasl link and cut tremor (Gtaa as in figure (3-3) as follows:

**Picture containing a table

Description created automatically**

**Figure (3-3): The confusion matrix of algorithms(SVM, NB, KNN)used to classify the link and cutting hums**

**Third: Results**

We have used a support machine algorithm Vector (SVM )Naif Khawarzmiyah bes (NB ) And a algorithm. Neighbor Closest (KNN )classification you whispered Connecting and cutting (In The beginning of the word.) And he take place Choose this Algorithms Based on several factors, most notably suitable size group Data with this Algorithms And a lot. characteristics Adopted On it in Classification process, take place work model training For all these Algorithms It has become clear from during Calculate the accuracy of the model for each algorithm that exceeds the support machine algorithm Vector (SVM ) Where i got the highest accuracy rate of the model reached %92, which is high and sufficient. لSolve the search problem BasicAnd it was. Basic in Access AlgorithmOn this high percentage is the large number of characteristics Used in classification (Done Use 8 Variables Independent) Which contributed in make model The rating is more realistic..

**Fourth: Research outputs and the benefits ofresearch:**

The aim of this completed project was to build a smart model that categorizes the link and cutting (at the beginning of the word) by using data classification algorithms in order to establish accurate and correct criteria in accurate writing Arabic texts to contribute and help adapt the technology to measure the quality of the most famous classification algorithms in distinction to serve the Arabic language, as well as the goal of research between the link and cutting whispers at the beginning of the word, and this intelligent model that was designed can contribute to the development of the most famous classification algorithms in the service of the Arabic language, as well as the goal of research between the two link and cutting whispers at the beginning of the word, and this intelligent model that was designed can contribute to the development of the most famous classification algorithms in the service of the Arabic language. Arabic asfollows:

* The new model can be used to review scientific research and ensure that the whisper is drawn in words correctly, contributing to the completeness of the elements of scientific research.
* The new model can be used to review news articles published on various social media by installing it in the browser plugins section to make sure that the whisper is drawn correctly in all the words of the published articles, which contributes to the development of the new Arabmedia.
* The new model can be used with tour applications and considered as an authentic part of mobile operating systems to make sure that translated commands are written in Arabic and started with a whispercorrectly.
* The new model can be matched with translation systems used in conferences, aircraft and trains, helping to display the humming properly.

Bottom line, we can use the new model as a software tool that can be combined with all software and technical systems that display readable or written Arabic text using software systems to categorize the whisper in words that are started with a link orcut.

**Recommendations:**

1. Apply algorithms suitable for dealing with purposes with multiple classification categories using indexing algorithms other than those that improvise purposes with binary categoriesonly.
2. Develop sklearn libraries to fully support Arabic.
3. Develop the model to act as a work for arabic audiotexts.

**Margins:**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

1. Dr. Alaa Mustafa Al-Hillis: Exploring Opinions in Arabic Comparative Sentences, Arab International Journal of Informatics, Volume 2, Issue 4, 2013.
2. Dr. Rana Zuhair Abdul Ghani Al Obaidi, Dr. Gida Abdul Aziz Al-Talib: A Comparative Study of Algorithms for Exploration of Opinions and Analysis of Emotions and Their Applications, Rafidain Journal of Computer Science and Mathematics Volume (12,) Issue II, 2018.
3. Amal Kazem Mirra ,,Thar Yer Jassim Kataa. 2019.) Applicationsof artificial intelligence in education from the point of view of universityteaching.
4. Amal Kazem Meera & Editing by Jassim Kataa. )2019.) Applications of artificial intelligence in education from the point of view of universityteaching.
5. Dr. M. Mustafa Obeid, (2018.) Advanced Analysis and Data Exploration, Arab Thought House, First Edition, Volume 1,Cairo.
6. Burkov, A. (2019). The hundred-page machine learning book (Vol. 1). Canada: Andriy Burkov.
7. Florin, G. (2011). Data mining: concepts, models and techniques. Springer-Verlag. Berlin Heidelberg.
8. Ian, H. W and Eibe, F. (2005). Data Mining: practical machine learning tools and techniques, Second Edition. Elsevier Inc. San Francisco: USA
9. Jiawei, H., Micheline, K. and Jian P. (2012). Data mining: concepts and techniques, Third edition. Elsevier Inc: USA.
10. Kalyani, G. and Jaya, A. Lakshmi Performance assessment of different classification techniques for intrusion detection. Journal of Computer Engineering (IOSRJCE).
11. Kalyani, G. and Jaya, A. Lakshmi Performance assessment of different classification techniques for intrusion detection. Journal of Computer Engineering (IOSRJCE).
12. Michael, J.A. B. and Gordon, S. L. (2004). Data mining techniques for marketing, sales, and customer relationship management, Second edition. Wiley Publishing, Inc. Indianapolis, Indiana: USA.
13. Amal Kazem Meera & Editing by Jassim Kataa. (2019.) Applications of artificial intelligence in education from the point of view of universityteachers.
14. Dr. M. Mustafa Obeid, (2018.) Advanced Analysis and Data Exploration, Arab Thought House, First Edition, Volume 1,Cairo.
15. The English statistician Thomas(lived during the period) 1761-1701AD(. . . drafted a special case of the famoustheory bearing his name, theorem ' Bayes) although it was not published in his life but published after his death by Richard Price.
16. Dr. M. Mustafa Obeid, (2018) Advanced Analysis and Data Exploration, Arab Thought House, First Edition, Volume 1,Cairo.
17. Amal Kazem Meera &Editing by Jassim Kataa. ) 2019.) Applicationsof artificial intelligence in education from the point of view of universityteaching.
18. Mohamed ben Makram bin Ali, Abu al-Fadl, Jamal al-Din, son of the African al-Ansari perspective. )1993.) 18 Tongue of the Arabs. Dar Sader - Beirut Edition: 3rd (1/17).
19. Mohammed bin Abdullah, son of Malik al-Taie al-Jiani, Abu Abdullah, Jamal al-Din. Investigation: Abdel Moneim Ahmed Haredi. (DT). Um al-Qura University Center for Scientific Research and Revival of Islamic Heritage Faculty of Sharia and Islamic Studies Mecca Edition: First (3/1466). (
20. Abu al-Fath Osman bin Juni al-Mosuli, Investigation: Faez Fares. (DT) shine in Arabic. Cultural Book House. Kuwait-

(226-225).

1. Mohammed Rafik Momen Al-Shobaki. (2015) Golden pearls in the explanation of the island introduction. Gaza- Palestine Edition: First (77).
2. Abdul Ahl Mohammed Al-Nacrates, Investigation: Mohammed Khalil Hras. (2003) Comprehensive in Arabic. National Book House - Lebanon Edition: First (178-162).
3. Taher Ben Mohamed Zouawi al-Birini of Algeria, Sheikh. (2019) . Scientific Books House - Beirut -Lebanon.

**Sources and references**

First: Arab sources and references

Picture containing text, document, screenshot

Description created automatically

Picture containing text, document, screenshot

Description created automatically

Second: Foreign sources and references

[13] Burkov, A. (2019). The hundred-page machine learning book (Vol. 1). Canada: Andriy Burkov. [14] Florin, G. (2011). Data mining: concepts, models and techniques. Springer-Verlag. Berlin Heidelberg.

[15] Ian, H. W and Eibe, F. (2005). Data Mining: practical machine learning tools and techniques, Second Edition. Elsevier Inc. San Francisco: USA.

[16] Jiawei, H., Micheline, K. and Jian P. (2012). Data mining: concepts and techniques, Third edition. Elsevier Inc: USA.

[17] Kalyani, G. and Jaya, A. Lakshmi Performance assessment of different classification techniques for intrusion detection. Journal of Computer Engineering (IOSRJCE).

[18] Michael, J.A. B. and Gordon, S. L. (2004). Data mining techniques for marketing, sales, and customer relationship management, Second edition. Wiley Publishing, Inc. Indianapolis, Indiana: USA.

**Supplement (A)**

N-spelled classification of novice words with a link or cut

Picture containing text

Description created automatically

