Paper Title\* (use style: paper title)

\*Note: Sub-titles are not captured in Xplore and should not be used

Miruna-Maria Cilibia  
Polytechnic University of TimișoaraTimișoara, România  
miruna.cilibia@gmail.com

line 1: 4th Given Name Surname  
line 2: *dept. name of organization*  
*(of Affiliation)*  
line 3: *name of organization   
(of Affiliation)*line 4: City, Country  
line 5: email address or ORCIDline 1: 2nd Given Name Surname  
line 2: *dept. name of organization   
(of Affiliation)*  
line 3: *name of organization   
(of Affiliation)*line 4: City, Country  
line 5: email address or ORCID

line 1: 5th Given Name Surname  
line 2: *dept. name of organization   
(of Affiliation)*  
line 3: *name of organization   
(of Affiliation)*line 4: City, Country  
line 5: email address or ORCIDline 1: 3rd Given Name Surname  
line 2: *dept. name of organization   
(of Affiliation)*  
line 3: *name of organization   
(of Affiliation)*line 4: City, Country  
line 5: email address or ORCID

line 1: 6th Given Name Surname  
line 2: *dept. name of organization   
(of Affiliation)*  
line 3: *name of organization   
(of Affiliation)*line 4: City, Country  
line 5: email address or ORCID

*Abstract*—This electronic document is a “live” template and already defines the components of your paper [title, text, heads, etc.] in its style sheet. *\*CRITICAL: Do Not Use Symbols, Special Characters, Footnotes, or Math in Paper Title or Abstract*. (*Abstract*)

Keywords—component, formatting, style, styling, insert (key words)

# Introduction (*Heading 1*)

This template, modified in MS Word 2007 and saved as a “Word 97-2003 Document” for the PC, provides authors with most of the formatting specifications needed for preparing electronic versions of their papers. All standard paper components have been specified for three reasons: (1) ease of use when formatting individual papers, (2) automatic compliance to electronic requirements that facilitate the concurrent or later production of electronic products, and (3) conformity of style throughout a conference proceedings. Margins, column widths, line spacing, and type styles are built-in; examples of the type styles are provided throughout this document and are identified in italic type, within parentheses, following the example. Some components, such as multi-leveled equations, graphics, and tables are not prescribed, although the various table text styles are provided. The formatter will need to create these components, incorporating the applicable criteria that follow.

# K-Means

K-means clustering is a method of vector quantization, originally from signal processing, that aims to partition n elements into k clusters in which each element belongs to the cluster with the nearest mean (cluster centers or cluster centroid), serving as a prototype of the cluster [1].

The intuition behind K-means is an iterative procedure that starts by guessing the initial centroids, and then refines this guess by repeatedly assigning examples to their closest centroids and then recomputing the centroids based on the assignments.

In the “cluster assignment” phase of the K-means algorithm, the algorithm assigns every training example to its closest centroid, given the current positions of centroids. Specifically, for every example we set

where is the index of the centroid that is closest to and is the position (value) of the centroid.

Given assignments of every point to a centroid, the second phase of the algorithm recomputes, for each centroid, the mean of the points that were assigned to it. Specifically, for every centroid we set:

where is the set of examples that are assigned to centroid .

# Fuzzy C-Means

Before you begin to format your paper, first write and save the content as a separate text file. Complete all content and organizational editing before formatting. Please note sections A-D below for more information on proofreading, spelling and grammar.

Keep your text and graphic files separate until after the text has been formatted and styled. Do not use hard tabs, and limit use of hard returns to only one return at the end of a paragraph. Do not add any kind of pagination anywhere in the paper. Do not number text heads-the template will do that for you.

## Abbreviations and Acronyms

Define abbreviations and acronyms the first time they are used in the text, even after they have been defined in the abstract. Abbreviations such as IEEE, SI, MKS, CGS, sc, dc, and rms do not have to be defined. Do not use abbreviations in the title or heads unless they are unavoidable.

## Units

* Use either SI (MKS) or CGS as primary units. (SI units are encouraged.) English units may be used as secondary units (in parentheses). An exception would be the use of English units as identifiers in trade, such as “3.5-inch disk drive”.
* Avoid combining SI and CGS units, such as current in amperes and magnetic field in oersteds. This often leads to confusion because equations do not balance dimensionally. If you must use mixed units, clearly state the units for each quantity that you use in an equation.
* Do not mix complete spellings and abbreviations of units: “Wb/m2” or “webers per square meter”, not “webers/m2”. Spell out units when they appear in text: “. . . a few henries”, not “. . . a few H”.

Identify applicable funding agency here. If none, delete this text box.

* Use a zero before decimal points: “0.25”, not “.25”. Use “cm3”, not “cc”. (*bullet list*)

## Equations

The equations are an exception to the prescribed specifications of this template. You will need to determine whether or not your equation should be typed using either the Times New Roman or the Symbol font (please no other font). To create multileveled equations, it may be necessary to treat the equation as a graphic and insert it into the text after your paper is styled.

Number equations consecutively. Equation numbers, within parentheses, are to position flush right, as in (1), using a right tab stop. To make your equations more compact, you may use the solidus ( / ), the exp function, or appropriate exponents. Italicize Roman symbols for quantities and variables, but not Greek symbols. Use a long dash rather than a hyphen for a minus sign. Punctuate equations with commas or periods when they are part of a sentence, as in:

*a**b* 

Note that the equation is centered using a center tab stop. Be sure that the symbols in your equation have been defined before or immediately following the equation. Use “(1)”, not “Eq. (1)” or “equation (1)”, except at the beginning of a sentence: “Equation (1) is . . .”

## Some Common Mistakes

* The word “data” is plural, not singular.
* The subscript for the permeability of vacuum **0, and other common scientific constants, is zero with subscript formatting, not a lowercase letter “o”.
* In American English, commas, semicolons, periods, question and exclamation marks are located within quotation marks only when a complete thought or name is cited, such as a title or full quotation. When quotation marks are used, instead of a bold or italic typeface, to highlight a word or phrase, punctuation should appear outside of the quotation marks. A parenthetical phrase or statement at the end of a sentence is punctuated outside of the closing parenthesis (like this). (A parenthetical sentence is punctuated within the parentheses.)
* A graph within a graph is an “inset”, not an “insert”. The word alternatively is preferred to the word “alternately” (unless you really mean something that alternates).
* Do not use the word “essentially” to mean “approximately” or “effectively”.
* In your paper title, if the words “that uses” can accurately replace the word “using”, capitalize the “u”; if not, keep using lower-cased.
* Be aware of the different meanings of the homophones “affect” and “effect”, “complement” and “compliment”, “discreet” and “discrete”, “principal” and “principle”.
* Do not confuse “imply” and “infer”.
* The prefix “non” is not a word; it should be joined to the word it modifies, usually without a hyphen.
* There is no period after the “et” in the Latin abbreviation “et al.”.
* The abbreviation “i.e.” means “that is”, and the abbreviation “e.g.” means “for example”.

An excellent style manual for science writers is [7].

# Implementation

The implementation of this project was done in Python.

// se mai poate extinde aici putin

## K-Means

The implementation for the K-Means algorithm consists of the following functions:

1. **kMeansInitCentroids(X, K):** Initializes K centroids utilized in the K-Means algorithm on the dataset X. The function selects random examples as initial centroids.
2. **findClosestCentroids(X, centroids):** Assigns each example in dataset X to the nearest centroid based on Euclidean distance calculation.
3. **computeCentroids(X, idx, K):** Computes new centroids by calculating the mean of data points assigned to each centroid.
4. **runkMeans(X, centroids, findClosestCentroids, computeCentroids, max\_iters=10):** Executes the K-Means algorithm by iteratively assigning examples to centroids and updating centroids until convergence or for a maximum number of iterations.
5. **runkMeansUnknownK(X, max\_k\_value=20, max\_iters=200):** Evaluates different cluster numbers (K) and selects the optimal number by measuring silhouette scores [2]. Utilizes the above functions for initializing centroids and clustering.

The silhouette score is a metric used to assess the quality of clustering in a dataset. It measures how well-separated clusters are, considering both the distance between data points within a cluster and the distance between different clusters. By using this metric, we can choose the best possible value for the

number of clusters, even when dealing with a dataset that has an unknown number classes.

## Fuzzy C-Means

## Plotting the data

The initial data, as well as the clusters resulted by the algorithms were displayed using the Seaborn library.

Visualizing K-means clusters using a **pairplot** offers a concise yet insightful representation of how the algorithm grouped data points. Leveraging Pandas for data organization and Seaborn for visualization, the **pairplot** showcases relationships between features and highlights distinct clusters through color-coded data points. The plotted centroids within the **pairplot** mark cluster centers, aiding in understanding cluster dispersion and feature relationships within each cluster.

# Results

## K-Means

1. Iris dataset

We tested our algorithm on the iris-150 dataset. After running the algorithm for 3 clusters and 200 maximum iterations, the results were the following:

We evaluated our algorithm using the silhouette score and had a result of 0.5528. For this score, a score close to 1 indicates that the clusters are well-separated, distinct, and have a good level of homogeneity. A score around 0 suggests overlapping clusters or clusters that are not well-separated. A score closer to -1 indicates possible incorrect clustering, where data points might have been assigned to the wrong clusters.

1. Dataset2

For the second dataset, we didn’t know the number of clusters beforehand, like we did with the iris dataset. Thus, we used the function that performs the algorithm for K = [2, …, 20] and 200 iterations and compared the silhouette score for each of them. The best number of clusters resulted was 2, with a silhouette score of 0.6545.

1. Our dataset

## Fuzzy C-Means

##### References

1. Hamerly, Greg, and Charles Elkan. "Learning the k in k-means." Advances in neural information processing systems 16 (2003).J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.
2. Wang, Fei, et al. "An analysis of the application of simplified silhouette to the evaluation of k-means clustering validity." Machine Learning and Data Mining in Pattern Recognition: 13th International Conference, MLDM 2017, New York, NY, USA, July 15-20, 2017, Proceedings 13. Springer International Publishing, 2017.K. Elissa, “Title of paper if known,” unpublished.
3. R. Nicole, “Title of paper with only first word capitalized,” J. Name Stand. Abbrev., in press.
4. Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, “Electron spectroscopy studies on magneto-optical media and plastic substrate interface,” IEEE Transl. J. Magn. Japan, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].
5. M. Young, The Technical Writer’s Handbook. Mill Valley, CA: University Science, 198