A Comparitive Analysis of Optimizing Classification Techniques on Fruits Pictures Dataset

Youssef George Fouad   
Faculty of Engineering  
Ain Shams UniversityCairo, Egypt  
email: 19p9824@eng.asu.edu.eg

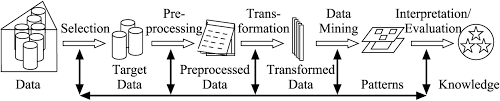
Kerollos Wageeh Youssef   
Faculty of Engineering  
Ain Shams UniversityCairo, Egypt  
email: 19p3468@eng.asu.edu.eg Jannah Ayman Amir  
Faculty of Engineering  
Ain Shams UniversityCairo, Egypt  
email: 19p1728@eng.asu.edu.eg

*Abstract*— Data mining, also known as data dredging or data archeology, provides several techniques to extract new and interpretable information from given datasets. In this paper, we compare the performance of several existing classification techniques on pictures of 3 fruits (pineapple, cocos, avocado) from fruits-360, an open-source dataset. In addition, we will compare how principal component analysis, one of the most famous dimensionality reduction techniques, affects the performance of the classification tasks.

Keywords— Knowledge Discovery in Database (KDD), Principal component analysis (PCA), Support Vector Machine (SVM), Radial Basis Function (RBF), K-Nearest Neighbors (KNN), Decision Trees, Image Processing, Classification Techniques.

# INTRODUCTION

Data mining is mainly used to transform raw data, that has no meaning in themselves, into useful, interpretable, and meaningful information. For instance, knowledge discovery in database (KDD) gives us methods and techniques to discover useful patterns and information from our raw data. The following figure gives an overview of the basic stages of the KDD process starting from the raw data, to selecting the target data we would work on, before preprocessing and transforming it to start the mining process that leads us to discovering some patterns, models, and useful information. This discovered information shall finally be evaluated and assessed before being used in business or further research use cases.



# PROBLEM SPECIFICATION

## Selecting a Template (Heading 2)

First, confirm that you have the correct template for your paper size. This template has been tailored for output on the A4 paper size. If you are using US letter-sized paper, please close this file and download the Microsoft Word, Letter file.

## Maintaining the Integrity of the Specifications

The template is used to format your paper and style the text. All margins, column widths, line spaces, and text fonts are prescribed; please do not alter them. You may note peculiarities. For example, the head margin in this template measures proportionately more than is customary. This measurement and others are deliberate, using specifications that anticipate your paper as one part of the entire proceedings, and not as an independent document. Please do not revise any of the current designations.

# DATASET

Having a dataset of high quality is essential for building a good classification model, also known as a classifier, so working on this paper we used fruit-360 dataset. The chosen dataset consists of hundreds of pictures taken for each of the 180 fruits available from different angles besides being split into training and test sets for training and testing purposes that will be discussed later in the paper.

Working on this comparative analytical paper, we decided to work on images of 3 fruits only, pineapple, cocos, and avocado. Our choice of these three fruits in specific gives us the opportunity to test the classification techniques in different ways as pineapples look like cocos in terms of color grades and cocos look like avocados in terms of the oval-like geometrical shape.

# evaluation criteria

In order to

## Authors and Affiliations

**The template is designed for, but not limited to, six authors.** A minimum of one author is required for all conference articles. Author names should be listed starting from left to right and then moving down to the next line. This is the author sequence that will be used in future citations and by indexing services. Names should not be listed in columns nor group by affiliation. Please keep your affiliations as succinct as possible (for example, do not differentiate among departments of the same organization).

### For papers with more than six authors: Add author names horizontally, moving to a third row if needed for more than 8 authors.

### For papers with less than six authors: To change the default, adjust the template as follows.

#### Selection: Highlight all author and affiliation lines.

#### Change number of columns: Select the Columns icon from the MS Word Standard toolbar and then select the correct number of columns from the selection palette.

#### Deletion: Delete the author and affiliation lines for the extra authors

## Identify the Headings

Headings, or heads, are organizational devices that guide the reader through your paper. There are two types: component heads and text heads.

Component heads identify the different components of your paper and are not topically subordinate to each other. Examples include Acknowledgments and References and, for these, the correct style to use is “Heading 5”. Use “figure caption” for your Figure captions, and “table head” for your table title. Run-in heads, such as “Abstract”, will require you to apply a style (in this case, italic) in addition to the style provided by the drop down menu to differentiate the head from the text.

Text heads organize the topics on a relational, hierarchical basis. For example, the paper title is the primary text head because all subsequent material relates and elaborates on this one topic. If there are two or more sub-topics, the next level head (uppercase Roman numerals) should be used and, conversely, if there are not at least two sub-topics, then no subheads should be introduced. Styles named “Heading 1”, “Heading 2”, “Heading 3”, and “Heading 4” are prescribed.

# approach

## Figures and Tables

#### Positioning Figures and Tables: Place figures and tables at the top and bottom of columns. Avoid placing them in the middle of columns. Large figures and tables may span across both columns. Figure captions should be below the figures; table heads should appear above the tables. Insert figures and tables after they are cited in the text. Use the abbreviation “Fig. 1”, even at the beginning of a sentence.

1. Table Type Styles

| Table Head | Table Column Head | | |
| --- | --- | --- | --- |
| Table column subhead | Subhead | Subhead |
| copy | More table copya |  |  |

1. Sample of a Table footnote. (*Table footnote*)
2. Example of a figure caption. (*figure caption*)

Figure Labels: Use 8 point Times New Roman for Figure labels. Use words rather than symbols or abbreviations when writing Figure axis labels to avoid confusing the reader. As an example, write the quantity “Magnetization”, or “Magnetization, M”, not just “M”. If including units in the label, present them within parentheses. Do not label axes only with units. In the example, write “Magnetization (A/m)” or “Magnetization {A[m(1)]}”, not just “A/m”. Do not label axes with a ratio of quantities and units. For example, write “Temperature (K)”, not “Temperature/K”.

# results and analysis

# conclusion

##### Acknowledgment *(Heading 5)*

We thank Dr. Nourhan Mohamed Zayed (Ain Shams University) and Eng. Mahmoud Soheil (Ain Shams University) for guidance and assistance in understanding essential data mining techniques.

##### References

The template will number citations consecutively within brackets [1]. The sentence punctuation follows the bracket [2]. Refer simply to the reference number, as in [3]—do not use “Ref. [3]” or “reference [3]” except at the beginning of a sentence: “Reference [3] was the first ...”

Number footnotes separately in superscripts. Place the actual footnote at the bottom of the column in which it was cited. Do not put footnotes in the abstract or reference list. Use letters for table footnotes.

Unless there are six authors or more give all authors’ names; do not use “et al.”. Papers that have not been published, even if they have been submitted for publication, should be cited as “unpublished” [4]. Papers that have been accepted for publication should be cited as “in press” [5]. Capitalize only the first word in a paper title, except for proper nouns and element symbols.

For papers published in translation journals, please give the English citation first, followed by the original foreign-language citation [6].

1. G. Eason, B. Noble, and I. N. Sneddon, “On certain integrals of Lipschitz-Hankel type involving products of Bessel functions,” Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955. *(references)*
2. J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.
3. I. S. Jacobs and C. P. Bean, “Fine particles, thin films and exchange anisotropy,” in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.
4. K. Elissa, “Title of paper if known,” unpublished.
5. R. Nicole, “Title of paper with only first word capitalized,” J. Name Stand. Abbrev., in press.
6. 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].
7. M. Young, The Technical Writer’s Handbook. Mill Valley, CA: University Science, 1989.