ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 4, Special Issue 19, April 2017

SAMPLE IDENTIFICATION OF ONION IN INDIA

Atul Mishra
SRM university ,
Ramapuram Campus Chennai ,India
mishra.atul775@gmail.com

Bhargava Krishna
SRM university ,
Ramapuram Campus Chennai ,India
bhargavakrishnapasupuleti@gmail.com

Anand Anil
SRM university ,
Ramapuram Campus Chennai ,India
anandanil@653gmail.com

S. Jaya kumar
Assit.professor
SRM university ,
Ramapuram Campus Chennai ,India

***Abstract*— One of the uses of Gas Chromatography-Mass Spectrometry (GC-MS) is in the detection of pesticide residues in fruit and vegetables. The following paper proposed the test on a Tomato using GC-MS technique and then handles that big data, using Data Mining. The system which we are proposing is the same technique applied over Indian Fruits and vegetables respectively and find out an Average result regarding the specific item chosen mainly Onions.**

The average value can be turned into a label just for the satisfaction of the users towards the items they tend to buy.

Different products can have very different permitted pesticide residue levels. For this reason, ensuring that the sample being analysed is– as labelled– is very important. Mislabelling can happen through human error during sample preparation in the laboratory, or by the client before being sent to the laboratory. A laboratory will often receive samples that have already been prepared, and the original material can not be visually identified. Mislabelling can have serious consequences, either in food safety, with produce being sent to market with pesticide levels above permitted thresholds, or economically, with produce being destroyed unnecessarily.

Keywords— GC-MS; sample identification; data mining; workflows

I. INTRODUCTION

Onion is an important vegetable crop in India and is an integral component of Indian culinary. Being an essential food item, it is also a highly politically sensitive commodity. This report analyzes the production, consumption, trade price behaviour of onion in India. Analysis of pesticide residues in fruit and vegetables is an important part of food safety monitoring. Guidelines and regulations are issued by government agencies, like the EU parliament or the Food and Drug Administration (FDA) in the USA, to protect consumers. Pesticide residues are often detected and quantified using Gas Chromatography-Mass Spectroscopy analysis (GC-MS), which we describe in more detail later. GC-MS typically has limits of detection from the sub parts-per-million down to the low parts-per-billion, depending on the type and setup of the instrument. Modern analysis involves searching for an ever increasing number of pesticide compounds, typically numbering in the hundreds.

Different products can have very different permitted pesticide residue levels. For this reason, ensuring that the sample being analysed is– as labelled– is very important. Mislabelling can happen through human error during sample preparation in the laboratory, or by the client before being sent to the laboratory. A laboratory will often receive samples that have already been prepared, and the original material cannot be visually identified. Mislabelling can have serious consequences, either in food safety, with produce being sent to market with pesticide levels above permitted thresholds, or economically, with produce being destroyed unnecessarily.

II. RELATED WORK

The question we examine in the base paper, is whether we can use the chromatographic data alone to distinguish between types of produce samples. Thus enabling us to identify whether a sample was labelled correctly and the appropriate pesticide level(s) applied.

The machine learning approach presented here uses a qualitative approach to GC-MS (i.e., determining sample type), in contrast to the usual quantitative approach in laboratories (i.e., determining compound concentrations).

A. Handling the Problem Statement

One way of answering this question is to determine a list of compounds that are expected to differentiate the various

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 4, Special Issue 19, April 2017

fruits and vegetables, and explicitly measure those in addition to the pesticides. Modelling would then be done on the determined concentration of these compounds. However, assuming that a suitable list of compounds could be readily found, this approach places a significant extra burden on the analysis (in terms of both time and cost). Per capita onion consumption has shown a significant growth in recent years. Based on NSSO data, per capita consumption at household level during 2004-05 to 2009-10 (5 years) for which data is available has increased by 32 percent in rural areas and 18.6 percent in urban areas, implying an average annual growth rate of 6.4 percent and 3.7 percent respectively (Table 2). On an annual basis per capita consumption in 2009-10 works out to be 9 kilograms in rural areas and 10 kilograms in urban areas. On a longer term, per capita monthly consumption increased from 380 grams per person in rural areas and 500 grams in urban areas in 1987-88 to 740 grams in rural areas and 850 grams in urban areas in 2009-10

III. Prepare Your Paper Before Styling

The paper is structured as follows: first, we give a short introduction into the GC-MS domain in Section 2, before explaining some of the pre-processing techniques in Section 3 along with some of the essential data regarding onion growth and consumption in India. Then, since we are using a workflow system for most of the tasks presented here, Section 5 explains the workflow engine in detail. After that, Section 6 shows the conducted experiments and their results, before Section 7 concludes the paper.

Before starting the paper let's have a quick view on some basics like architecture and the figures which will be playing role while the time of execution. [6] discussed about a method, Sensor network consists of low cost battery powered nodes which is limited in power. Hence power efficient methods are needed for data gathering and aggregation in order to achieve prolonged network life. However, there are several energy efficient routing protocols in the literature; quiet of them are centralized approaches, that is low energy conservation.

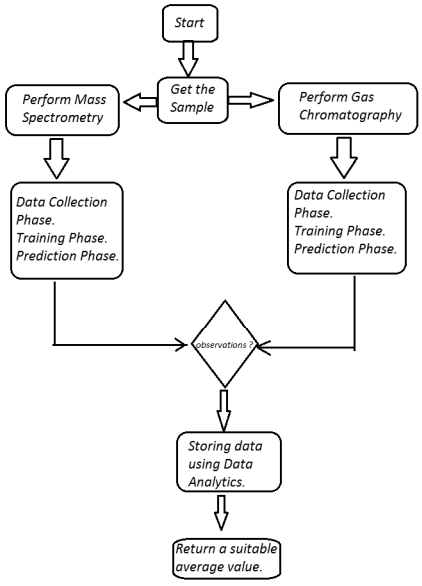

Figure 1

A. *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.

B. *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 units when they appear in text: "...a few henries," not "...a few H."
- Use a zero before decimal points: "0.25," not ".25." Use "cm3," not "cc." (*bullet list*)

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 4, Special Issue 19, April 2017

C. Algorithm and Pseudocode

- First we Initialise the values found by experimenting.
- The values are stored and applied using the Techniques of Data Mining.
- When a sample passes from the prototype, the laser scans the basic parameters such as density, salinity and then matches with the stored data.
- The sample must be in the range of it's parameters as to satisfy the requirements.
- Categorizing occurs on the bases of their satisfaction towards the requirements for passing.
- The sample passing n-2 or n-1 conditions out of are considered to be in the "A" Category, while the others are classified into "B" and "C".
- Pseudo code:

```
//importing parameters from Data Mining
//declaring variables such as parameters
//performing for mass and density only
from datamining *import
mass=int(input("Enter Mass:"))
density=int(input("Enter Density"))
for n in parameter.datamining():
    If mass = = parameter.datamining.mass
  and density==parameter.datamining.density
    return True
      else:
        return False
```

D. Some Common Mistakes

- The word "data" is plural, not singular.
- The subscript for the permeability of vacuum μ_0, and other common scientific constants, is zero with subscript formatting, not a lowercase letter "o."
- In American English, commas, semi-/colons, 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].

IV. Using the Template

After the text edit has been completed, the paper is ready for the template. Duplicate the template file by using the Save As command, and use the naming convention prescribed by your conference for the name of your paper. In this newly created file, highlight all of the contents and import your prepared text file. You are now ready to style your paper; use the scroll down window on the left of the MS Word Formatting toolbar.

A. Authors and Affiliations

The template is designed so that author affiliations are not repeated each time for multiple authors of the same affiliation. Please keep your affiliations as succinct as possible (for example, do not differentiate among departments of the same organization). This template was designed for two affiliations.

1) *For author/s of only one affiliation (Heading 3):* To change the default, adjust the template as follows.

a) Selection (Heading 4): Highlight all author and affiliation lines.

b) Change number of columns: Select the Columns icon from the MS Word Standard toolbar and then select "1 Column" from the selection palette.

c) Deletion: Delete the author and affiliation lines for the second affiliation.

2) *For author/s of more than two affiliations:* To change the default, adjust the template as follows.

a) Selection: Highlight all author and affiliation lines.

b) Change number of columns: Select the "Columns" icon from the MS Word Standard toolbar and then select "1 Column" from the selection palette.

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 4, Special Issue 19, April 2017

c) Highlight author and affiliation lines of affiliation 1 and copy this selection.

d) Formatting: Insert one hard return immediately after the last character of the last affiliation line. Then paste down the copy of affiliation 1. Repeat as necessary for each additional affiliation.

e) Reassign number of columns: Place your cursor to the right of the last character of the last affiliation line of an even numbered affiliation (e.g., if there are five affiliations, place your cursor at end of fourth affiliation). Drag the cursor up to highlight all of the above author and affiliation lines. Go to Column icon and select "2 Columns". If you have an odd number of affiliations, the final affiliation will be centered on the page; all previous will be in two columns.

B. 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.

C. Figures and Tables

1) 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.

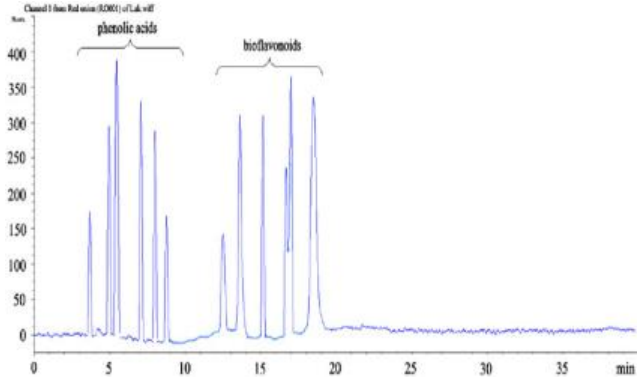

Figure 2

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."

We suggest that you use a text box to insert a graphic (which is ideally a 300 dpi resolution TIFF or EPS file with all fonts embedded) because this method is somewhat more stable than directly inserting a picture.

To have non-visible rules on your frame, use the MSWord "Format" pull-down menu, select Text Box > Colors and Lines to choose No Fill and No Line.

V. WORKFLOW

The Basic workflow can be understood by an Implementation. The proposing paper suggest to develop a prototype which has the algorithm to declare the onion fit for health or not.

- The parameters such as Salinity, weight per sample,density,acidity are obtained with values from the GC-MS application.
- To create the prototype the requirements are good knowledge of DataMining, well know individual with Python and Data Structures, Electronics, Scanning Ray to detect,etc. hardwares.
- Now the Sample is passed through the prototype to check the result. Now we take such a sample for which already we know the result. If the passed sample satisfies then it's a SUCCESS.
- Now we propose to categorize the onions on the basis to the satisfaction of result they give.

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 4, Special Issue 19, April 2017

- Suppose a sample satisfies 1 condition(assuming the loop with 3 Conditions), then it is Classed B Category. Now this category onions could be sent to the Restaurants. The basic motive of sending them there is to let them maintain their Hygiene to the best of their Ability.
- Category A Onions are the one which satisfies 2 or more conditions and they can be labelled as **"A"** and sent to the market.
- When labelled it provides a satisfaction to the end user. Also we propose this project to the Government as the Private Companies or Middle Users may do product frauds.

V. CURRENT PROPOSING SYSTEM

The already proposed system contains the GC- MS performed on Tomato samples. Now we are proposing this method on onions to identify the hygienic onions .Firstly we will get the samples of onions by performing mass-spectrometry gas chromatography, then store those values in the system now pass the onions through the scanning prototype. The proposed system invokes the required parameters before processing is initiated using the applications of Data Mining. The onion given as a sample to the system is analyzed and compared to the required parameters defined initially. The sample is accepted if the values obtained are met, else rejected If the parameters are not matched. We will categorize those onions into three A,B,C; category A consists of fresh and hygienic onions which can be consumed by people.as grade B consists good onions but less than category A onions as these are most probably used in hotels ,now C consists of waste onions. As grade C onions are not fit for human consumption, they are utilised in replanting the onions for the next harvest. This is an economical method to save funds for seeds.

VI. CONCLUSION

We have shown in this paper that data mining can be used to support analysts in laboratories in terms of quality control, minimizing the impact of accidental sample swaps. Clearly, proper pre-processing of the data is paramount for achieving good results, obtaining as few misclassifications as possible. The promising results of the "origin" test could lead to future work, involving how this approach can be applied, for example, for testing produce shipments for potential fraud. This will help in differentiating the onions and providing the right quality and quantity in a very cost effective manner.

Acknowledgment (HEADING 5)

The preferred spelling of the word "acknowledgment" in America is without an "e" after the "g." Avoid the stilted expression "one of us (R. B. G.) thanks ...". Instead, try "R. B. G. thanks...". Put sponsor acknowledgments in the unnumbered footnote on the first page.

References

[1] Hall, M., E. Frank, G. Holmes, B. Pfahringer, P. Reitman, and I. H. Witten. The weka data mining software: An update. SIGKDD Explorations, 11, 2009.
[2] Holmes, G., D. Fletcher, and P. Reutemann. Predicting polycyclic aromatic hydrocarbon concentrations in soil and water samples. In Proc International Congress on Environmental Modelling and Software, 2010.
[3] Lud¨ascher, B., I. Altintas, C. Berkley, D. Higgins, E. Jaeger, M. Jones, E. A. Lee, J. Tao,and Y. Zhao. Scientific workflow management and the kepler system. Concurrency and Computation: Practice and Experience, 18:1039–1065, 2006.
[4] Nadeau, C. and Y. Bengio. Inference for the generalization error. Machine Learning, 52(3):239–281, 2003.
[5] Mierswa, I., M. Wurst, R. Klinkenberg, M. Scholz, and T. Euler. Yale: Rapid prototyping for complex data mining tasks. In Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-06), 2006.
[6] Christo Ananth, S.Mathu Muhila, N.Priyadharshini, G.Sudha, P.Venkateswari, H.Vishali, "A New Energy Efficient Routing Scheme for Data Gathering ",International Journal Of Advanced Research Trends In Engineering And Technology (IJARTET), Vol. 2, Issue 10, October 2015), pp: 1-4
[7] WikiPedia. Gas chromatography-mass spectrometry—Wikipedia, the free encyclopedia, 2012. [Online; accessed 23-Feb-2012].
[8] WikiPedia. Gas chromatography-mass spectrometry—Wikipedia, the free encyclopedia, 2012. [Online; accessed 23-Feb-2012].