International Journal of Advanced Smart Convergence Vol.5 No.1 47-50 (2016)
http://dx.doi.org/10.7236/IJASC.2016.5.1.47

IJASC 16-1-7

Big data Analysis using Python in Agriculture Forestry and Fisheries

So hee Kim*, Min Soo Kang**, Yong Gyu Jung***,+

*,**,***,+ *Department of Medical IT Marketing, Eulji University, Korea*
thgml8411@hanmail.net, mskang@eulji.ac.kr, ygjung@eulji.ac.kr

Abstract

Big Data is coming rapidly in recent times and keep the vast amount of data was utilized them. These data are utilized in many fields in particular, based on the patient data in the medical field to increase the therapeutic effect, as well as re-incidence to better treatment, lowering the readmission rates increased the quality of life. In this paper it is practiced to report basis of the analysis and verification of data using python. And it can be analyzed the data through a simple formula, from Select reason of Python to how it used; by Press analysis of Agriculture, Forestry and Fisheries research. In this process, a simple formula can be used that expression for analyzing the actual data so it taking advantage of the use of functions in real life

Keywords: *Data Mining, Weka, Machine Learning, Classification, Python, Equation*

1. Introduction

Big Data is coming in recent times and keep the vast amount of data was utilized them. These data are utilized in many fields in particular, based on the patient data in the medical field to increase the therapeutic effect, as well as re-incidence to better treatment, lowering the readmission rates increased the quality of life. Data mining is a "To create a new and useful knowledge from large amounts of data", data processing, data summarization, machine learning, pattern recognition, visualization, statistics, knowledge extraction technique, etc requires the skills of a variety of fields. Machine Learning provides a methodology to extract the information from the source data in the database as the primary technical-based data mining. For example, if a model consisting of a parameter, on the basis of past experience or training data is referred to as a computer program or a training learning act to optimize the parameters of the model. The learned model can predict the results from the new data have never met in the learning process. In this paper, we compare the performance of the algorithm proceeds separation is made based on the ML. Undergo a process of using a known Supervised algorithm and to ensure that any algorithm J48 compared to the performance of this REPTree provide more accurate classification results in the goal of this study is to find a more accurate prediction algorithms.

Manuscript Received: Dec. 22, 2015 / Revised: Jan. 10, 2016/ Accepted: Feb. 3, 2016
+ Corresponding Author: ygjung@eulji.ac.kr
Tel: +82-31-740-7190
Dept. of Medical IT marketing, Eulji University, Korea

2. Related Research

In practice two things before the analysis of Agribusiness data. Both analysis and verification based on a 2013 survey value of Agriculture, Forestry and Fisheries through the python. Find out the annual increase or decrease the growth rate using the basic formula of arithmetic. In the first research, learn to farm population but with it is not to represent the overall distribution. So investigate the value of the attempt by farmers evaluate the distribution by city. Due to the above it can be seen that the data on which areas are agricultural workers. Also, I can briefly know the overall number of farmers and the distribution of the above <Table 1> and <Table 2> that value of the draft for who engaging in agriculture or wish to agriculture and business.

Table 1. Farm Households and Population (2009~2013)

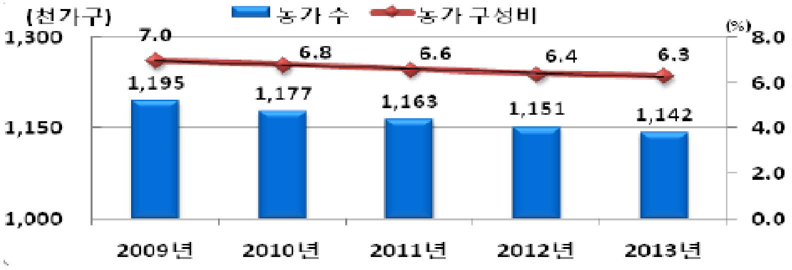
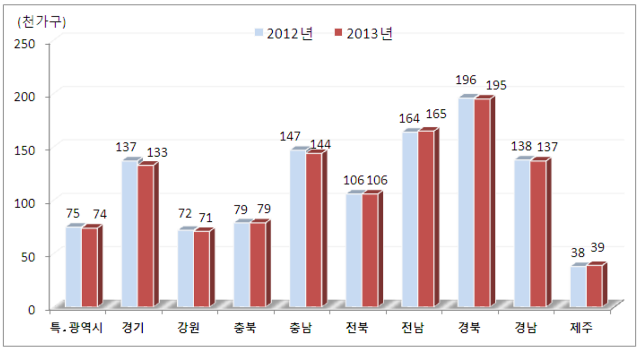


Table 2. farmers by Cities and provinces (2012~2013)



To make decimal value output in Python, it is shown as an integer by descending. So, show up decimal if you want to be a real numbers. It can be grouped as largely five methods. It can be seen that appears to be real number upset point even execution.

(1) Divide After entering the ".0" after the number of divided characters.
(2) Divide After entering the ".0" after the number to divided characters.
(3) To specify the float () values to the number to divided characters.
(4) To specify the float () values to the number to divided characters.
(5) Both characters can share with you to share and specify a float () value.

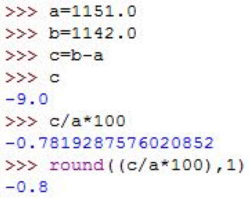
It can be rounded up to specific value so that the point can also be infinitely. When all the data can be seen that show up first place. In this case can be expressed by using the round function. When referring to the above formula, If you enter the round (1.12,1) then 1.1 comes out. And it show the number of first place by rounded in second place.

3. Simulation

As a tool for the study was developed using WEKA v3.6.10 at Waikato University, the data used is big data portal. The experiment described below was used for several methods of data mining and the data analysis showed a 10-fold value of the Cross-Validation. k-fold Cross Validation is a way to ensure that there is no unique set of one share, compared with those of 'k' other full.

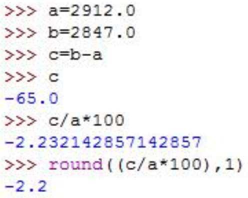
3.1 Farmers and farmer population practice

It can be obtained composition and growth rate through a simple formula in the same manner as in the below Figure 1 and Figure 2



```
>>> a=1151.0
>>> b=1142.0
>>> c=b-a
>>> c
-9.0
>>> c/a*100
-0.7819287576020852
>>> round((c/a*100),1)
-0.8
```

Figure 1. Farmhouse Results

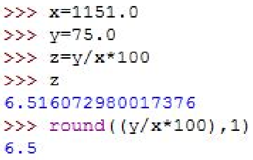


```
>>> a=2912.0
>>> b=2847.0
>>> c=b-a
>>> c
-65.0
>>> c/a*100
-2.232142857142857
>>> round((c/a*100),1)
-2.2
```

Figure 2. Farm Population Results

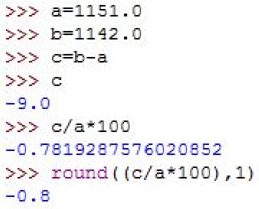
3.2 Practice by city farm

In the same manner, It is shown as in the below Figure 3 and Figure 4. It can be also obtained composition and growth rate through a simple formula.



```
>>> x=1151.0
>>> y=75.0
>>> z=y/x*100
>>> z
6.516072980017376
>>> round((y/x*100),1)
6.5
```

Figure 3. Composition Results



```
>>> a=1151.0
>>> b=1142.0
>>> c=b-a
>>> c
-9.0
>>> c/a*100
-0.7819287576020852
>>> round((c/a*100),1)
-0.8
```

Figure 4. sensitization, growth rate results

4. Conclusion

In this paper, it was practice focused why and how to use Python to select. Based on the data in the process looked data coming from the public analyzes actual data how used in the real world and how convenient for used. These data can be utilized in many fields in particular, based on the patient data in the medical field to increase the therapeutic effect, as well as re-incidence to better treatment, lowering the readmission rates increased the quality of life. It was analyzed the data through a simple formula, from Select reason of Python to how it used; by Press analysis of Agriculture, Forestry and Fisheries research. In this process, a simple formula can be used that expression for analyzing the actual data so it taking advantage of the use of functions in real life.

.

References

[1] Witten, Ian H., and Eibe Frank. Data Mining: Practical machine learning tools and techniques. Morgan Kaufmann, 2005.

[2] Baldi, Pierre, and Søren Brunak. Bioinformatics: the machine learning approach. MIT press, 2001.

[3] Kononenko, Igor. "Machine learning for medical diagnosis: history, state of the art and perspective." Artificial Intelligence in medicine 23.1 (2001): 89-109.

[4] Zhou, Zhi-Hua, and Min-Ling Zhang. "Solving multi-instance problems with classifier ensemble based on constructive clustering." Knowledge and Information Systems 11.2 (2007): 155-170.

[5] Bellazzi, Riccardo, and Blaz Zupan. "Predictive data mining in clinical medicine: current issues and guidelines." International journal of medical informatics 77.2 (2008): 81-97.

[6] Zhu, Xiaojin. "Semi-supervised learning literature survey." (2005).

[7] Cho, Sung-Bae, and Hong-Hee Won. "Machine learning in DNA microarray analysis for cancer classification." Proceedings of the First Asia-Pacific bioinformatics conference on Bioinformatics 2003-Volume 19. Australian Computer Society, Inc., 2003.

[8] Yongheng Zhao and Yanxia Zhang, Comparison of decision tree methods for finding active objects, Advances of Space Research, 2007

[9] D. L. Gupta, A. K. Malviya, Satyendra Singh Performance Analysis of Classification Tree Learning Algorithms, International Journal of Computer Applications (0975 – 8887) Volume 55– No.6, October 2012