

GLOBAL JOURNAL OF ENGINEERING SCIENCE AND RESEARCHES A STATISTICAL ANALYSIS ON THE ASPECTS OF PATTERN RECOGNITION

Aishwaryasri K^{*1} & Kukatlapalli Pradeep Kumar²

^{*1}BTech Student, Dept. of CSE, CHRIST (Deemed to be University), Bengaluru, India

²Asst.Professor, Dept. of CSE, CHRIST (Deemed to be University), Bengaluru, India

ABSTRACT

In this growing era of technology, there are still some parts of India where illiteracy is a major issue. Considering one part of the country, surveying in one city each from every direction, the rate of illiteracy is found and the SPSS, a statistical tool which assists in plotting large data with ease is used to perform analysis. This tool was acquired by IBM and is used to analyses, manipulate and generate graphs through the summarized data. The parameters are compared by performing two tests namely, Chi-square test and One-way ANOVA test. Lastly considering two parameters a graph is plotted against rate of illiteracy and cities. SPSS has wide applications in the field of data mining, data warehouse, artificial intelligence, big data analytics, machine learning etc.

Keywords: - SPSS tool, pattern recognition, illiteracy, chi-square test, One-way ANOVA test.

I. INTRODUCTION

Illiteracy being a major concern in India which is varied among different factors such as age, gender, region etc. This paper uses an already existing dataset from a verified government website consisting of different regions in Karnataka, India in each direction. The regions considered are Bengaluru (South), Belgaum (North), Udupi (Central Region), Shimoga(Southwestern). The age factor is mainly used to obtain illiteracy rate in male and female. The graphs shown in further sections are compared against illiteracy among age groups in each region. The Chi-square test is used to perform whether the null hypothesis is true or not considering the nominal values in the data given. It is used to perform to detect a significant difference between expected frequencies and the observed frequencies in multiple categories. The ANOVA test is carried out to find the difference between groups. Pattern recognition is the branch of machine learning that focuses on the regularities of data. It is also same as data mining which discovers patterns in large dataset.

II. LITERATURE STUDY

This section provides a brief understanding on the latest advancements of pattern recognition in Artificial Intelligence collected from newest research articles.

It is observed that the problem in which slideshow presentations is accessed at a proximity to operate the presentation. The usage of clicker when used is of batteries and an additional cost [1]. This paper proposes a solution to this problem by using the hand pattern which is recorded using a inbuilt camera. The tool used is MATLAB and it features are obtained using Principal Component Analysis (PCA). The algorithm used is K-Nearest Neighbor Algorithm (KNN). This designed system functioned with an accuracy of 90%. When experimented with different orientations of hand the accuracy ranged between 82.78% - 100% with the user 1.5 meters away from the projection. When the user moved away with a distance of 2 meters the accuracy was decreased by 7.22% i.e. to 75.56%. When slide consisting with texts was used to experiment there was a minimal difference and the accuracy ranged between 82.78% - 100% and the orientation angles varied from -60 to +60 degrees. The overall performance is depended on lighting condition of the area, the distance of the laptop to the user, and how the laptop is oriented toward the user. Skin detection function and addition of more patterns to the database was recommended for better performance.

The overall improvement of GHI (Global Horizontal Irradiance) with 1-hour ahead is considered using three sets of models for experimental purpose in which they are divided according to the previous analysis of data. [2] The

procedure follows as the first two data models forecast GHI for first four daylight and for the remaining hours are determined by weather classification model called optimal machine learning model through the third model set. The pattern recognition algorithm used is Support Vector Machine (SVM) . The weather pattern is recognized by the data collected in a single day with the ratio of actual GHI to clear sky GHI. The information concluded in this paper is that proportion of sunny days, partially cloudy days, and cloudy days in G3 are 71%,.

23%, and 6%, respectively. The GHIs were forecasted between 7am – 7pm. The developed framework showed better performance than the persistence model for 1 hour ahead GHI forecasting.

A solution to the Local Sequence Alignment which is also known as Smith -Waterman Algorithm which maintains optimal sequence alignment using R. Schutzhold's pattern recognition quantum algorithm. [3] Quantum technology is widely used in the field of bioinformatics which analyzes massive amounts of sequence data for genes and proteins. The steps of the proposed algorithm are reformed and powered by quantum mechanics computing theory. By the comparisons of sequences from the proposed algorithm a binary and unstructured dataset are formed to form relevant patterns. Spatial light modulator is used to achieve this. After the datasets are compared it is presented and results that local sequence algorithm is more efficient for large scale sequences. The feature selection problem is efficiently solved by QFT by obtaining relevant information from the datasets.

The data analytical framework for pattern recognition of unutilized data in a bulk transmission grid is done using clustering algorithm for operation of pattern recognition. [4] This is experimented on a real-grid and the results are seen.

A new speed-independent feedback index for effectively recognizing walking pattern for real-time control of a robotic suit is considered and the chosen index value remains the same for each walking patterns at different speeds. [5] It varies accordingly, from young to elderliness. Index value should be small for better performance of the device. This paper represents a survey which was experimented on elderly male and female, young male and female. Their walking speed and walking ratio is calculated using few derived formulas. Walk ratio and walking speed is considered for comparing and obtaining a pattern considering 6 individuals with respected to gender and age. The experimental result proves that speed-independent feedback index is almost constant for different walking patterns and walking speeds. The results recommend that the speed independent feedback index should be brought to down to a small value for maintaining a good walking pattern for elderly.

Spiking neural networks also abbreviated as SNNs to control animats known as artificial animals in a task which is experimented in a temporal pattern recognition and foraging 2D environment consisting of two objects namely target and distractor where target produces a temporal pattern consisting two components and distractor produces random patterns unlike the target and the patterns produced are a combination of target emitted components.[6] It is shown that the evolved animals are strong to the changes in the environment that is changes such as strength of the actuators, duration of signals, intervals between signals in the pattern and between patterns. The conclusion proves that there is less robustness changing the strength of the sensors, the intervals between signals, and increased duration of signals.

A digital electric meter reading recognition with respect to horizontal and binary patterns as it the basic process in the field of digital image processing was used to obtain meter readings in every possible climatic condition automatically. The main purpose was to is to find a method which automatically generates an electric bill through the device digital electric meter. [7] Hence, digital electric meter reading framework is proposed. Firstly, the selected image is normalized and converted into a YCbCr image. Since the device's region is resolute to light color, particularly green color, so the reading region is extracted from YCbCr image based on Cb and Cr. Secondly, Canny Edge operator is used to detect edge image i.e. it detects the points where the image brightness changes sharply and the individual digits' edges are filled up through morphological operation. The individual digits are segmented from an edge image through vertical projection. Next, the segmented digits are filled and thinned to detect shapes of the digits. Hence, from these digits and shapes the horizontal and vertical binary patterns are extracted. Through this

process different conditional digital meter images have been tested and most of them were achieved successfully with 94.10% recognition rate.

The Particle Swarm Organization (PSO) uses a technique called Computational Intelligence (CI) in a laboratory environment to observe the existence of the pattern between the net energy consumption by the electric loads in the building and the ambient temperature along with the occupancy state of the building. [8] The obtained pattern is used to predict the consumption of energy the near future. The electric loads considered are ventilation, lighting, heating and air conditioning (HVAC) units intelligent monitoring and control capabilities using internet of things (IoT) devices and other technologies. This process can reduce energy waste due to excess generation. A PSO algorithm is used to solve this using historic data containing ambient temperature, occupancy, and energy consumption gathered over a period of six days (July 23– 28, 2017) was presented.

A method to quantify the uncertainty of Pattern Recognition using Neural Networks is done in LOCAs in a nuclear reactor using Artificial Neural Networks. [9] The first method uses error estimation by a series association scheme and the second approach uses Bayesian model averaging technique.

Error Estimation by Series Association: As the name suggests, a series of ANNs are used to predict the uncertainty. A secondary ANN aids in predicting the error of the primary ANN. Although this yields lower accuracy, computation is rather easy. This approach provides a varying percentage of accuracy - between 44% to 63% with 5 networks.

On the other hand, Bayesian Model approach requires training a set of ANNs with the same structure which is an overhead on computation. Computation can be reduced using parallel computing techniques but requires additional resources. This approach ensures an accuracy equal to 100% when 50 networks were used.

Both the approaches have their own trade-offs.

III. RESULTS AND DISCUSSION

The below elicits about output screens and graphs related to a selected dataset. This is run in an IBM's statistical tool called as SPSS

The number of illiterate candidates is compared with age and the region using the SPSS tool by performing two tests namely

One-way ANOVA and Chi-square test. The two tests conclude in the following way; when ANOVA is performed with City as one variable and number of Illiterate candidates as another variable, the obtained significant value with City is 0.076 which is greater than 0.05 (the base probability value). This case proves that the null hypothesis to be accepted and, and the alternative hypothesis is rejected as it means that the mean of number of illiterate students among the cities is equal and there is no significant association with 'city' and 'number of illiterate candidates'. Secondly, when the ANOVA test is performed with age the sig value obtained is 0.010 which is lesser than 0.05 (the base probability value) proving that the null hypothesis is rejected and alternative accepted which means that the mean of number of illiterate candidates is not equal in regard with age. Hence, with age it proves that there is significant association with number of illiterate candidates. The table is shown in Table I.

The chi-square test when performed it is found that the expected count is less than 0.05 (the base probability value) and the minimum count is 0.25. The parameters considered are illiteracy of male and female with city. The Chi Square assumption is violated as it is more than 20 % (obtained 100%). Considering the Pearson Chi Square the significant value obtained is 0.321 which is accepting the null hypothesis and rejecting the alternative. This test proves that there is no significant association between the illiterate male and female candidates and the city. This table is shown in Table II. Finally, a bar graph is plotted against the age and number of illiterate candidates in each city considered for ease of detailed study. Fig 1 is the graphical representation of the city Bangalore. The graph

shows the age group between 0-6 has the highest number of illiterate candidates of 484982 in number and the age group of 11 has the least number of illiterate candidates of 3477. Fig 2 is the graphical representation of the city Belgaum. This graph shows the age group 0-6 as the highest number of illiterate candidates of 47642 and the age group of 11 has the least number of illiterate candidates with 160 in number. Fig 3 is the graphical representation of the city Shimoga. This city also the highest number of illiterate candidates in the age group of 0-6 with 33128 in number and the least of 143 in age group of 11. Fig 4

TABLE I. ANOVA RESULTS TEST

		Sum of Squares	df	Mean Square	F	Sig.
City	Between Groups	135.833	103	1.299	4.455	.076
	Within Groups	1.167	4	.292		
	Total	135.000	107			
Age	Between Groups	57942.185	103	562.545	13.805	.010
	Within Groups	163.000	4	40.750		
	Total	58105.185	107			

is the graphical representation of the city Udupi. It shows that the highest number of illiteracy candidates range between the age group of 0-6 with 18799 in number and the least in the age group of 12. Comparing all the four graphs, we conclude that in these regions majority of the children start their education at the age of 11. Fig 5 represents the architecture of the SPSS tool.

TABLE II. CHI-SQUARE RESULTS

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	320.000 ^a	309	.321
Likelihood Ratio	296.667	309	.683
Linear-by-Linear Association	5.445	1	.020
N of Valid Cases	108		

416 cells (100.0%) have expected count less than 5. The minimum expected count is .25.

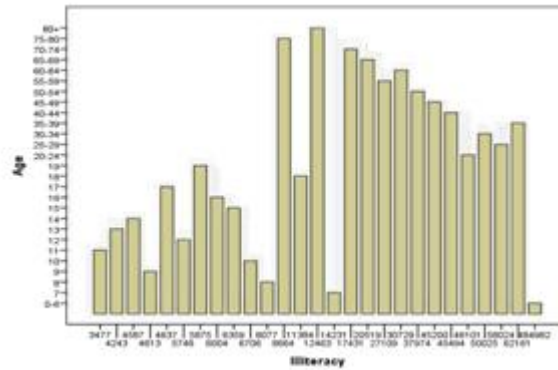


Fig.1 Bar Graph representation for Bangalore city on 'age' and 'illiteracy'.

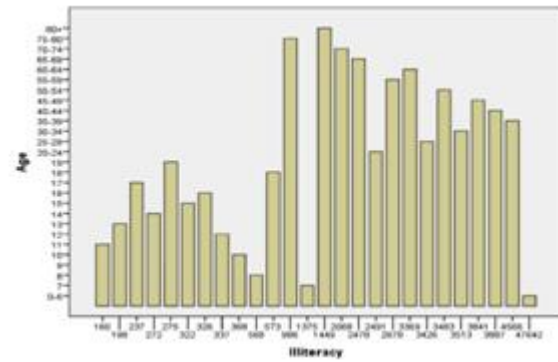


Fig.2 Bar Graph representation for Belgaum city on 'age' and 'illiteracy'.

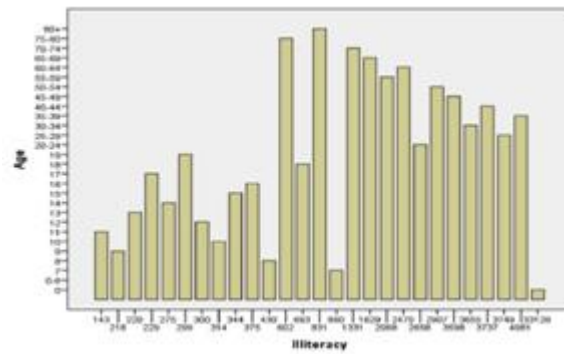


Fig.4 Bar Graph representation for Udipi city on 'age' and 'illiteracy'.

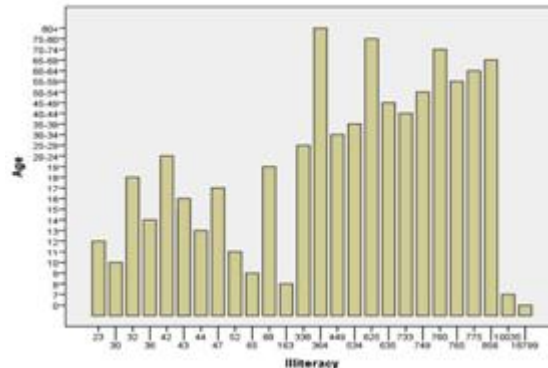


Fig.3 Bar Graph representation for Shimoga city on 'age' and 'illiteracy'.

IV. CONCLUSION

A base level research is carried out with an experimental analysis on a dataset with variables as age, city and illiteracy. The statistical measures namely, ANOVA test and Chi-Square test were carried out on the basis of testing the null hypothesis. Graphical illustrations were represented taking the above three variables into account. This analysis is performed as a part of internal term report making on outcome based academic activity.

V. ACKNOWLEDGMENT

It gives me immense pleasure and satisfaction in submitting this term report on Pattern Recognition using SPSS tool. In the endeavor of preparing this term report many people gave me a helping hand. So, it is my duty and pleasure to express my deep regards. The data set collected is from the URL data.gov.in.

On the outset of this term paper I express my sincere regards to my guide Mr. Kukatlappalli Pradeep Kumar (CSE department) who gave me an opportunity to work on this topic and made me present my views on topic through term paper. I also acknowledge my gratitude to the infrastructure provided by dept. of CSE, CHRIST (Deemed to be University) in carrying this experimental analysis.

REFERENCES

- Jonathan S. Del Mundo¹, Eric Martin R. Lingal, Ralph F. Los Baños, Vero M. Romero Jr., Jehiel D. Santos, Seigfred V. Prado, Edison A. Roxas "Slideshow Presentation Control through Hand Pattern Recognition using Web Camera" 978-1-5386-0912-5/17 ©2017 IEEE.
- Cong Feng, Mingjian Cui, Meredith Lee, and Jie Zhang, Bri-Mathias Hodge Siyuan Lu, Hendrik F. Hamann "Short-term Global Horizontal Irradiance Forecasting Based on Sky Imaging and Pattern Recognition" 978-1-5386-2212-4/17 ©2017 IEEE
- Konstantinos Prousalis, Nikos Konofaos "Quantum Pattern Recognition for Local Sequence Alignment" 978-1-5386-3920-7/17/ ©2017 IEEE
- Xianzhuang Liu, Wei Hu, Le Zheng, Yong Min, Xialing Xu, Yong LI, Rui Yu," Operation Pattern Recognition via Mass in Bulk" 978-1-5386-3920-7/17/ ©2017 IEEE.
- Ru Ma, Junqiang Li, Shanhai Jin, Shijie Guo, Kazunobu Hashimoto, Shijie Dai "A speed-independent feedback index for walking pattern recognition for a walking assistive robotic suit" 2017 IEEE 8th International Conference on CIS & RAM, Ningbo, China
- Chama Bensmail, Volker Steuber and Neil Davey, Borys Wr'obel "Evolving Spiking Neural Networks to Control Animats for Temporal Pattern Recognition and Foraging" 978-1-5386-2726-6/17/©2017 IEEE
- Atif Anis¹, Md. Khaliluzzaman^{1,2}, Mohammad Yakub¹, Niloy Chakraborty¹, and Kaushik Deb^{2*} "Digital Electric Meter Reading Recognition Based on Horizontal and Vertical Binary Pattern" 2017 3rd

International Conference on Electrical Information and Communication Technology (EICT), 7-9 December 2017, Khulna, Bangladesh

8. *Guneet Bedil, Ganesh Kumar Venayagamoorthy, Rajendra Singh “Pattern Recognition for Electric Energy Consumption Prediction in a Laboratory Environment” 978-1-5386-2726-6/17©2017 IEEE*
9. *Silvia Tolo, T. V. Santhosh, Gopika Vinod, Uchenna Oparaji, Edoardo Patelli “Uncertainty Quantification Methods for Neural Networks Pattern Recognition” 978-1-5386-2726-n6/17/ ©2017 IEEE.*