International Journal of Engineering &Technology, 7 (3.12) (2018) 925 -926

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

Linear Genetic Programming, Naïve Bayes algorithm - Their Applications in Geotechnical Engineering

Vishweshwaran M*

SASTRA Deemed University, Thanjavur, Tamil Nadu
**Corresponding Author Email: vishweshwaran@civil.sastra.edu*

Abstract

Modeling soil and rock pose challenges due to uncertainties in their complex behavior. In the present study, linear genetic programming and Naïve Bayes are used in classification of liquefied and non-liquefied data. Soil and seismic parameters influencing the soil liquefaction potential are used to develop the models. Genetic Programming is the automatic creation of computer programs to perform a selected task using Darwinian natural selection. Linear genetic programming forms a peculiar subset of genetic programming where computer programs in a population are constituted as successive repetition of instructions from imperative programming language. Naive Bayes methods are supervised learning algorithms by applying Bayes' theorem with the "naive" assumption of independence among all the sets of the features. Accuracy of results of classification for linear genetic programming, Naïve Bayes were found to be 94.12% and 90.59% respectively.

1. Introduction

Inputs of datasets were adopted from literature of Goh. All the models were trained and tested. The inputs used in the development of model are as follows: earthquake moment magnitude denoted as M_w, maximum horizontal acceleration indicated by a_{max}, total stress σ_v, effective stress σ'_v sleeve friction ratio R_f and cone tip resistance (q_c). Output Liquefaction was represented as 'Yes or No' in Naïve Bayes algorithms respectively and '1 or 0' in linear genetic programming.

2. Modeling

Linear Genetic Programming

Linear genetic programming investigates the development of imperative computer programs written as linear sequences of instructions. The process of one run takes place as follows. The first step involves creating a random population of computer programs followed by random selection of four of the programs. Determining the fitness function and two best programs out of the four remain in the population and reproduce. Operators such as mutation and crossovers are used by repetition of the process until convergence conditions are fulfilled. In the output, 1 indicates Liquefied case and 0 indicates non liquefied case. Testing was performed for 33% of the data after training and validation.
The selected best program is shown in the figure –

```
L0: f[0]+=-1.3605182170867925;
L1: f[0]/=Input005;
L2: f[1]+=f[0];
L3: f[0]+=Input004;
L4: f[0]*=Input005;
L5: f[0]-=Input001;
L6: f[0]+=Input004;
L7: f[1]/=f[0];
L8: f[0]/=f[0];
L9: f[0]+=Input004;
L10: f[1]/=f[0];
L11: f[0]-=f[0];
L12: f[0]+=Input004;
L13: f[0]*=Input005;
L14: f[0]-=Input001;
L15: f[0]+=Input004;
L16: f[1]/=f[0];
L17: f[0]+=f[0];
L18: f[0]+=f[1];
L19: f[0]+=Input004;
L20: f[0]-=-0.4943120479583745f;
L21: f[0]-=Input000;
L22: f[0]*=0.13169598579406745f;
L23: f[0]+=Input004;
L24: f[0]-=-0.4943120479583745f;
L25: f[0]+=Input004;
```

Fig. 1: Selected best program using Linear genetic programming

Copyright © 2018 Authors. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

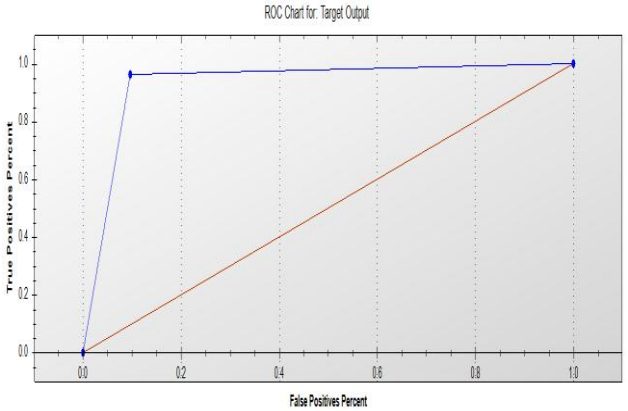

Fig. 2: ROC chart for the target output using Linear Genetic Programming

Naïve Bayes Algorithm

Bayes theorem is applied in Naïve Bayes algorithm. In spite of its simplicity, Naive Bayes is effective in classification. Bayes' Theorem offers a way that we can find the probability of a hypothesis given our past knowledge. P(m|n) is the probability of data m rendered that the hypothesis n was true. P(n) is the probability of hypothesis n being true irrespective of the data). This is known as prior probability of n. P(m) is the probability of the data (irrespective of the hypothesis). Instead of attempting to calculate the values of each attribute value P(m1, m2, m3|n), they are presumed to be conditionally independent given the target value and found as P(m1|n) * P(m2|n) and so on. Testing of the model was performed for 33% of the available data.

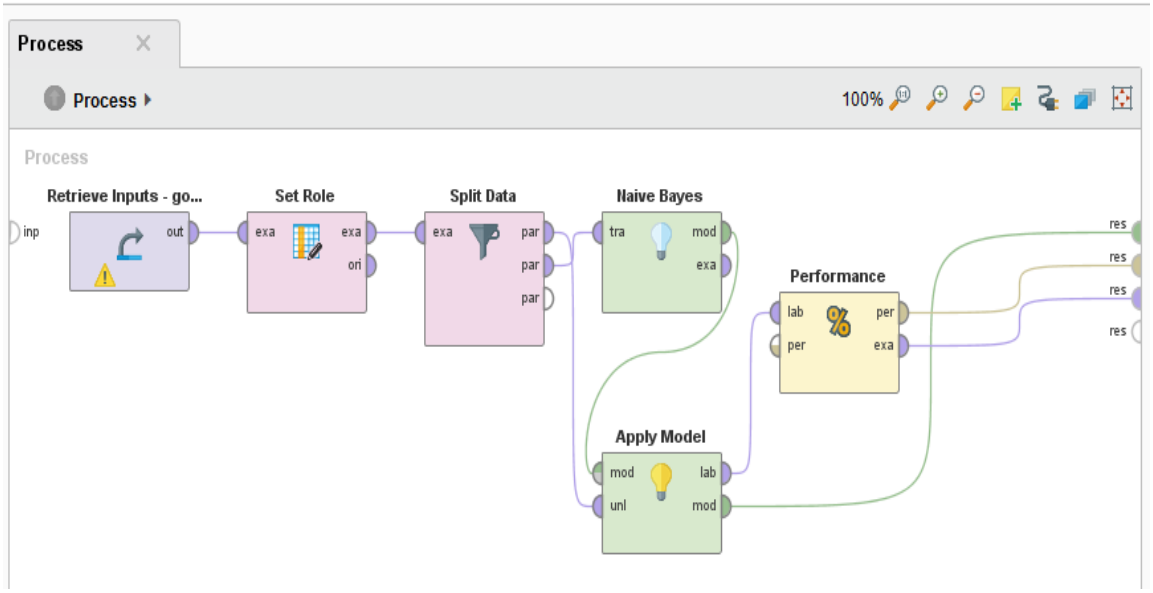

Fig. 3: Naïve Bayes process and parameters

3. Results and Discussion

Linear Genetic Programming

Confusion matrix was found to be:

Table 1: Confusion matrix

	True yes	True no
Predicted yes	52	3
Predicted no	2	28

Accuracy of the program was 94.12%.
The accuracy of Linear Genetic Programming was better than the Naïve Bayes algorithm for the chosen datasets.

Naïve Bayes algorithm

Table 2: Class precision and class recall

	True yes	True no
Predicted yes	51	4
Predicted no	4	26

Accuracy of the model was 90.59%.
Other statistical parameters are also calculated and represented in the table.

Table 3: Statistical parameters

Statistical parameters	Linear Genetic Programming	Naïve Bayes model
Accuracy (%)	94.12	90.59
Sensitivity (%)	96.3	92.73
False positive rate (%)	9.68	13.33
Specificity (%)	90.32	86.7

4. Conclusion

Machine learning provides reasonable to excellent results for the Liquefaction data. They have high potential applications in various Geotechnical engineering linear and non linear analyses. In this analysis, 2 models were developed using linear genetic programming and Naïve Bayes algorithm. The accuracy of linear genetic programming is better than Naïve Bayes model.

References

[1] Anthony T.C. Goh, S.H. Goh, "Support vector machines: Their use in geotechnical engineering as illustrated using seismic liquefaction data" Computers and Geotechnics. 34 (2007) 410–421.
[2] Amir Hossein Gandomi, Amir Hossein Alavi, "A new multi-gene genetic programming approach to non-linear system modeling. Part II: geotechnical and earthquake engineering problems". Neural Computing and Applications. (2012) 21:189–201
[3] Efstratios F. Georgopoulos et al. "Genetic Programming Modeling and Complexity Analysis of the Magnetoencephalogram of Epileptic Patients". Information Systems Development. 2009. 40. 383-391.
[4] Frank D. Francone et al. "Discrimination of Unexploded Ordnance from Clutter Using Linear Genetic Programming". Genetic Programming Theory and Practice III. 2006. 49-64.