Chapter 8

Result and Conclusion

Result

As we have applied four different algorithms on the Credit card dataset taken from Kaggle. On Some calculations these Algorithms provides Accuracy, Precision, and Recall values of the Implementation process, Which are tabulated as :

KNN Algorithm

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Fraud Rate** | **Accuracy** | **Precision** | **Sensitivity** | **Specificity** | **AUC** |
| **25%** | 0.6262 | 0.6000 | 0.5959 | 0.6527 | 0.6240 |
| **50%** | 0.6311 | 0.6159 | 0.5894 | 0.6687 | 0.6300 |
| **75%** | 0.6564 | 0.6481 | 0.6222 | 0.6879 | 0.6560 |
| **100%** | 0.6862 | 0.6893 | 0.6349 | 0.7339 | 0.6870 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Fraud Rate** | **Accuracy** | **Precision** | **Sensitivity** | **Specificity** | **AUC** |
| **25%** | 0.9617 | 0.9726 | 0.9467 | 0.9755 | 0.9610 |
| **50%** | 0.9686 | 0.9636 | 0.9700 | 0.9674 | 0.9690 |
| **75%** | 0.9658 | 0.9467 | 0.9816 | 0.9523 | 0.9670 |
| **100%** | 0.9612 | 0.9474 | 0.9713 | 0.9522 | 0.9620 |

SVM Algorithm

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Fraud Rate** | **Accuracy** | **Precision** | **Sensitivity** | **Specificity** | **AUC** |
| **25%** | 0.9585 | 0.9650 | 0.9452 | 0.9701 | 0.9590 |
| **50%** | 0.9827 | 0.9932 | 0.9702 | 0.9940 | 0.9830 |
| **75%** | 0.9755 | 0.9977 | 0.9511 | 0.9979 | 0.9770 |
| **100%** | 0.9715 | 0.9914 | 0.9498 | 0.9924 | 0.9730 |

Random Forest Algorithm

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Fraud Rate** | **Accuracy** | **Precision** | **Sensitivity** | **Specificity** | **AUC** |
| **25%** | 0.9521 | 0.9281 | 0.9810 | 0.9226 | 0.9520 |
| **50%** | 0.9686 | 0.9881 | 0.9539 | 0.9862 | 0.9700 |
| **75%** | 0.9370 | 0.9651 | 0.9180 | 0.9600 | 0.9390 |
| **100%** | 0.9509 | 0.9771 | 0.9315 | 0.9740 | 0.9530 |

Logistic Regression Algorithm

Conclusion

Finally, as the results in table shows that highest accuracy in given by the Random forest algorithm when applied all algorithms with SMOTE data balancing technique. These results are formed on taking different fraud data rates in the dataset befor applying to the algorithms.

Chapter 9

Future Scope

* More Algorithms can be applied on the same dataset to get good results.
* Combination of two or more algorithms of data mining are very helpful in determining the best results. This combination of algorithms is known as Hybrid Algorithm.

* Also by making classifiers in present algorithms and applying new upcoming algorithms may be helpful for improvement.
* Choosing only those attributes which most affect the liver if slight change in quantity occurs in body. And applying Different approaches may improve the outcomes.

Chapter 10

References

1. Masoumeh Zareapoor, P.Shamsolmoali{2015}. Application of Credit Card Fraud Detection: Based on Bagging Ensemmble Classifier.International Conference on Intelligent Computing And Convergence(ICCC-2015).
2. S. Kotsiantis, D. Kanellopoulos, P. Pintelas (2006). Handling imbalanced datasets: A review. International Transactions on Computer Science and Engineering.
3. G.H. John, P. Langley (1995). Estimating continuous distributions in Bayesian classiﬁers. in: Proceedings of the 11th Conference on Uncertainty in Artificial Intelligence, (1995); 338 — 345.
4. I3I R.J. Bolton, D.J. Hand (2001). Unsupervised profiling methods for fraud detection. In Conference on credit scoring and credit control, Edinburgh.
5. D. Kibler. D.W. Aha, M. Albeit (1989). Instance-based prediction of real-valued attributes. Computational Intelligent
6. P.K. Chan, W. Fan, A.L. Prodromidis. S.J. Stolfo ( 1999). Distributed Data Mining in Credit Card Fraud Detection. IEEE Intelligent Systems. pp 67—74.
7. C. Cortes, V. Vapnik (1995). Support vector networks. Machine Learning. 20:273—297.
8. T.M. Cover, P.E. Hart (1967). Nearest neighbor pattern classification. IEEE Trans. Information Theory, l3(1):21—27.
9. G. Potamilis (2013). Design and Implementation of a Fraud Detection Expert System using Ontology-Based Techniques. A dissertation submitted to the University of Manchester for the degree of Master of Science in the Faculty of Engineering and Physical Sciences.
10. E. David (2012). Bayesian inference-the future of online fraud protection. Computer Fraud & Security, 8-11.
11. S. Ghosh, D.L. Reilly. (1994). Credit Card Fraud Detection with a Neural- Network. In Proceedings of the International Conference on System Science, pages 621-630.
12. Jha. Sanjeev, G. Montserrat, J.C. Westland (2012). Employing transaction aggregation strategy to detect credit card fraud. Expert system with application, 39: 12650-12657.
13. L. Breiman - Random forests. Machine Learning. (2001). Vol (45); 5—32.
14. J. Piotr., AM. Niall. J.D. Hand, C. Whitrow, J. David (2008). Off the peg and bespoke classiﬁers for fraud detection. Computationa Statistics and Data Analysis, 52
15. L. Qibei & J. Chunhua (2011). Research on Credit Card Fraud Detection Model Based on Class Weighted Support Vector Machine. Journal of Convergence Information Technology, 6(1). 62-68.
16. S. Maes, K. Tuyls, B.Vanschoenwinkel, B.Manderick (1993). Credit card fraud detection using Bayesian and neural networks. In Proceedings for the First International NAISO Congress on Nettro Fuzzy Technologies, pages 261-270.
17. E.W.T. Ngai, H.Yong., Y.H.Wong, Y.Chen, X. Sun (2011). The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature. Decision Support Systems, 50:559—569
18. M. Zareapoor, Seeja, K.R. & M. Alam, Afshar (2012). Analyzing Credit Card Fraud Detection Techniques Based On Certain Design Criteria. International Journal of Computer Application. 52(3):35-42.
19. E.Aleskerov,B.Freisleben, B.Rao, CARDWATCH: a neural network based database mining system for credit card fraud detection, in computational intelligence for ﬁnancial engineering, Proceedings of the IEEE/IAFE, IEEE, Piscataway, NJ, 1998, pp. 220–226.
20. M. Artis, M. Ayuso, M. Guillen(2002), Detection of automobile insurance fraud with discrete choice models and misclassiﬁed claims, The Journal of Risk and Insurance 69 (3}325–340.
21. R.J. Bolton, D.J. Hand{2001}, Unsupervised proﬁling methods for fraud detection, Conference on Credit Scoring and Credit Control, Edinburgh.
22. R.J. Bolton, D.J. Hand (2002), Statistical fraud detection: a review, Statistical Science 17 (3) 235–249.
23. R. Brause, T. Langsdorf, M. Hepp {1999}, Neural data mining for credit card fraud detection, Proceedings of the 11th IEEE International Conference on Tools with Artiﬁcial Intelligence.
24. C.R. Bollinger, M.H. David (1997), Modeling discrete choice with response error: food stamp participation, Journal of the American Statistical Association .
25. R. Caruana, N. Karampatziakis, A. Yessenalina{2008}, An Empirical Evaluation of Supervised Learning in High Dimensions, in: Proceedings of the 25th international Conference on Machine Learning Helsinki, Finland, July,.
26. R. Caruana, A. Niculescu-Mizil{ 2006}, An Empirical Comparison of Supervised Learning Algorithms, in: Proceedings of the 23rd International Conference on Machine Learning, Pittsburgh, Pennsylvania, June.
27. S.B.Caudill, M.Ayuso, M.Guillen{1999}, Fraud detection using a multinomial logit model with missing information, The Journal of Risk and Insurance.
28. P.K. Chan, W. Fan, A.L. Prodromidis, S.J. Stolfo, Distributed Data Mining in Credit Card Fraud Detection, Data Mining, (November/December).