|  |  |
| --- | --- |
| Paper | Description |
| Sarah S. Alduayj, Kashif Rajpoot, *Predicting Employee Attrition using Machine Learning* | The developing interest for ML among business pioneers and leaders requires that scientists investigate its utilization inside business associations. One of the significant issues confronting business pioneers inside organizations is the dropping of skilled workers. This paper analyzes employee attrition with ML algorithms. Utilizing data made by I.B.M Watson, three exercises were carried out for predicting employee attrition. The first exercise included training the actual imbalanced data using these ML models: support vector machine (SVM), random forest classifier and K-nearest neighbor (KNN). The second exercise concentrated on utilizing ADASYN to deal with class imbalance, and then re-training on the new data utilizing the previously mentioned ML algorithms. The third exercise included utilizing data under sampling to balance the classes. Accordingly, training an ADASYN-adjusted data with KNN (K = 3) accomplished the best results, getting high F-measure scores. |
| D. G. Gardner, L. V. Dyne and J. L. Pierce, *The effects of pay level on organization‐based self‐esteem and performance: A field study,* 2004 The British Psychological Society | Enterprise managements presume that higher pay grades will maintain and increase effective operation of employees in the future. This paper inspects the effect of high pay grade as incentive for improved employee performance. The authors collected data from various sources in three stages in a span of 12 months, and evaluated the effects of pay grade on ensuing performance and self-esteem. They hypothesized that pay grade effects on employees’ self-esteem brought about improvement in performance. The hypothesis is based on the assumption that pay levels in a company conveys how much the company appreciates the efforts of the individual and thus affects employees’ self-esteem and hence performs better. |
| V. V. Saradhi, G. K. Palshikar, *Employee churn prediction,* Expert Systems with Applications  Volume 38, Issue 3 (2011) | Customer churn is a serious issue for most businesses, as loss of customer affects profits and bringing in new customers is not easy. Prediction models for customer churn can be beneficial in developing customer retention programs. Employee attrition has a similar effect on businesses, causing operational disturbances, customer discontentment and effort and time wasted in finding replacements. This paper surveys and compares some machine learning algorithms that have been employed to design predictive customer churn models. The authors also carried out a case study designing and comparing predictive employee attrition models. They also propose a value model that can identify how many of the attrition affected employees were valuable. |
| Ioannis Paparrizos, B. Barla Cambazoglu, Aristides Gionis, *Machine Learned Job Recommendation*, Proceedings of the fifth ACM conference on Recommender systems, 2011 | In this paper, the authors design a supervised machine learning-based job recommendation system. This algorithm utilizes all previous job changes and data linked with organizations and employees to predict next job transition of an employee. They trained a machine learning model using large dataset of job transitions of approximately 5 million employees publicly available on the Internet. The data on each employee is divided into three sections: the first section contains personal information, second section contains professional background of the employee and third section contains educational background of the employee. Experiments conducted by the authors have proved that job transitions can be predicted accurately. The machine learning algorithm used is a decision tree + naive Bayes hybrid classifier (DTNB). |
| Juanjuan Wang, Mantao Xu, Hui Wang, Jiwu Zhang, *Classification of Imbalanced Data by Using the SMOTE Algorithm and Locally Linear Embedding*, 2006 8th international Conference on Signal Processing | Imbalanced data classification often occurs in few important  practical applications such as data mining and patter recognition in medical sciences. Most of the current classification techniques are designed by presuming the training data used is distributed in an even manner. Yet, this data has a critical bias issue when training dataset is greatly imbalanced which leads to below par performance. SMOTE is a major approach of oversampling the positive class or the minority class. Yet, this data is restricted to a presumption, that the local space in the middle of any two positive cases is positive or is a part of the minority class, that does not always be correct in the situation when the training dataset is non-linearly separable. Yet, plotting the training dataset into a more linearly separable space can fix this issue. In this paper, the authors have combined Locally Linear Embedding algorithm (LLE) and SMOTE so that oversampling can be done on datasets that are non-linearly separable. Experiments have shown that this technique yields better results than traditional SMOTE. |
| Andrew Christian Flores, Rogelyn I. Icoy, Christine F. Peña, Ken D. Gorro, *An Evaluation of SVM and Naive Bayes with SMOTE*  *on Sentiment Analysis Data Set*, 2018 International Conference on Engineering, Applied Sciences, and Technology | Sentiment Analysis is an assessment of language that can ascertain a person's sentiments and is usually used as data for ML. In this paper, the authors do a study of Support Vector Machine (SVM) algorithm: Sequential Minimal Optimization (SMO) with SMOTE and Naive Bayes Multinomial (NBM) technique with SMOTE for categorization of data with the Sentiment Analysis data collected by students of San Carlos University. SMO is a technique used to work out quadratic  programming problem in training SVMs. A GUI called Weka using a suite of ML techniques for data mining, is utilised to pre-process and classify the data. They were able to conclude that SMOTE was effective depending on the way the data was processed prior to performing SMOTE and the kind of training and testing is furthermore a way of obtaining reliable outcomes. They also concluded that oversampling might not enhance noisy sentiment analysis data which does not have meaning. |
| E. Moncarz, J. Zhao and C. Kay, *AN EXPLORATORY STUDY OF US LODGING PROPERTIES' ORGANIZATIONAL PRACTICES ON EMPLOYEE TURNOVER AND RETENTION* | The motivation behind this study is to explore US lodging properties' employee preservation schemes, and to look at the effect of those practices on worker turnover and preservation. With directories of Hotels and Lodging properties, a helpful specimen data of 24 administration organizations were chosen. A self-administered mail study tool was created to gauge and test schemes on employee turnover and preservation. Utilizing SPSS 16.0, two measurable tests are utilized to study hypotheses. Correlation analysis is utilized to recognize the connections among predictor and response factors. In the same manner, regression analysis is utilized to analyze the connections among predictor and response factors hypothesizing the efficacy of carrying out the human resource management schemes on management and non-management preservation and turnover will vary. The discoveries uncover that Job Culture and professional and hiring practices impact employee preservation. Besides, Corporate Goals and leadership, and Work Commendation and Rewards were found to decidedly decrease employee turnover. Due to the survey technique and the sparse responses, speculation of the investigation discoveries is restricted. Further replication of this study is suggested. The discoveries could prepare lodging associations and industry experts with the instruments to decrease worker turnover and to retain more employees. This should positively affect productivity. |
| Alao D. & Adeyemo A. B., *ANALYZING EMPLOYEE ATTRITION USING DECISION TREE ALGORITHMS*, Computing, Information Systems & Development Informatics Vol. 4 No. 1 March, 2013 | Employee turnover is a genuine worry in information based associations. When workers depart from a company, they take priceless implicit information which is advantageous in the industry. For a company to consistently have a better edge against its competition, it should limit employee attrition. This paper recognizes worker’s traits which add to the prediction of worker attrition in companies. Data of a few hundreds of workers from Nigeria were utilized for the investigation. The population structure and occupation associated data of the worker were the primary information that was utilized to arrange the worker into some predetermined attrition classes. WEKA and See5 for Windows were utilized for producing decision tree models and rules. The outcomes were then utilized for building up a predictive model which was utilized to anticipate new instances of employee attrition. A structure to make a program tool which can execute the guidelines created in this paper was additionally suggested. |
| Piyasak Jeatrakul, Kok Wai Wong, and Chun Che Fung,  *Classification of Imbalanced Data by Combining the Complementary Neural Network and SMOTE Algorithm* | In recent times, research teams have discovered that an imbalanced dataset might be one of the hindrances to a lot of machine learning techniques. While learning ML algorithms, whenever the proportion of minority classes and majority classes was notably unalike, algorithms will in general be commanded by the majority classes while the attributes of the minority classes are barely recognized. Subsequently, the classification precision of the minority classes might be lower than the classification precision of majority classes. The attributes in the minority classes are ordinarily hard for recognizing completely. In this study, to stabilize the distribution, the joining of two strategies, Complementary Neural Network (CMTNN) and SMOTE, are proposed. CMTNN was used as an under-sampling procedure, whereas, SMOTE is utilized as an over-sampling method. CMTNN is utilized because of being able to predict "Truth" data and additionally "False" data as well. SMOTE is used since it can make new cases instead of repeating the already existing cases. |
| He-Yong Wang,  *Combination approach of SMOTE and biased-SVM for Imbalanced datasets* | Imbalanced data learning is risky as conventional ML approaches fail to give satisfactory outcomes because of skewed class distribution. Rather than the two usual solutions to this problem, undersampling and oversampling, a new approach to develop the classifiers from imbalanced datasets is proposed in this paper by joining SMOTE and BiasedSVM approaches. Often, real-world data chiefly consists of regular data along with a few examples of anomalous data. The expense of miscategorizing such deviant data into a regular model is frequently a worse than that of the converse mistake. Test results affirms that the proposed mix approach of SMOTE and biased SVM can accomplish better classifier performance. |
| Kung-Jeng Wanga, Bunjira Makonda, Kun-Huang Chena, Kung-Min Wang,  *A hybrid classifier combining SMOTE with PSO to estimate 5- year survivability of breast cancer patients* | Data processing issues are difficult in health services because of huge, complicated, diversified, hierarchical time series data. The yearly number of deaths brought about by cancers is around million globally and breast cancer is one of the deadliest forms of cancer. To know how long a patient can survive and to alleviate them in order to make a decision with respect to therapy and financial arrangements is crucial. In the interim, wrong categorization can result in wasted finances and/or wrong therapy to cure it. In this study, the authors propose new algorithms to enhance the efficacy of classification for 5-year survivability of breast cancer patients from a huge imbalanced dataset. Results from this indicate that the hybrid algorithm of SMOTE + PSO + C5 is the best one for predicting among all the combinations. They deduced that, executing SMOTE in suitable searching algorithms such as PSO and classifiers such as C5 can remarkably enhances the efficacy of grouping for classification for huge imbalanced datasets. |
| S. Kaur and R. Vijay,  *HISTOPATHOLOGICAL IMAGE ANALYSIS: A REVIEW* | The Employee turnover has consistently been an important matter of worry for organizations. In the present period of globalization there are plentiful opportunities for talented individuals in this world, therefore, workers always move from one organization to another. Due to this organizations are facing the issue of employee attrition. A huge degree of worker turnover is profoundly damaging to both the association and the employees. The most effective method to decrease employee attrition is a definitive test for HR executives. This article presents a comprehensive perspective of attrition and retention of workers in this competitive scenario regarding Retail Industry. Alongside other factors, Job Satisfaction has been considered as a significant source of attrition and retention. The research is based on their literature review and also from the data accessible on the web. |
| Akhil Gokuldas Warrier, Rajiv Prasad, *Motivators, Hygiene Factors and Job Satisfaction of Employees in IT Sector in India* | The IT field is a significant benefactor for the Indian economy throughout the past few decades. However, lately, numerous new opportunities are opening up for the best talents. Subsequently, employee attrition is very high in the IT segment nowadays. In this study the authors look at the role that Herzberg's factors on hygiene and motivation have in guaranteeing job contentment of the workers in this industry. It has been found to work with certain distinctions in various nations, particularly Asia, due to social contrasts. There has barely been made any effort to comprehend the manner in which the factors pointed out by Herzberg influence job contentment of workers within the IT field. For this paper, the authors inspected the role of these two factors. Information was gathered from 153 IT professionals. It was discovered that in spite of what was anticipated from the results, the cleanliness attribute assumes a more grounded part in forecasting job contentment of the workers. The ramifications of this result are talked about in this study. |
| Tince Etlin Tallo, Aina Musdholifah,  *The Implementation of Genetic Algorithm in Smote (Synthetic Minority Oversampling Technique) for Handling Imbalanced Dataset Problem* | Imbalanced data can influence the performance of standard classifier algorithms that causes biased outcomes towards majority classes. The SMOTE technique fixes the imbalanced data issue by making synthetic cases of minority classes. Yet, the usage of SMOTE brought about over-generalization on the grounds that synthesized instances have a similar amount no matter what the distribution of instances is. Accordingly, the borders between classes are vague. The SMOTE-Simple Genetic Algorithm (SMOTESGA) strategy is utilized to decide the sampling rate of every example so as to get unequal number of synthesized cases. The trials were executed with some imbalanced data by comparing the classification results calculated with G-means and F-Measure scores. The outcomes of the use of genetic algorithm along with SMOTE could enhance the classification outcome by acquiring better scores. |
| A. Al-Radaideh and E. A. Nagi,  *USING DATA MINING TECHNIQUES TO BUILD A CLASSIFICATION MODEL FOR PREDICTING EMPLOYEES PERFORMANCE* | Manpower and human resources is a worrying issue for organizations' administration because their highest interest is in employing the profoundly skilled staff who are relied upon to perform exceptionally. In this paper, data mining strategies were used to design a classification technique to predict the performance of workers. For designing the classification technique the CRISP-DM data mining strategy was embraced. Decision tree was the chief data mining technique utilized for constructing the classification technique, where quite a few classification rules were created. For approving the produced model, many trials were performed utilizing genuine information gathered from several organizations. The technique is planned to be utilized for predicting new candidates' performance. |