Study on Predicting Employee Turnover using HR Data

B Sai Ashish[1]

Abstract

HR(Human Resources) plays a major role in predicting the employee turnover with is most important in managing the role of which Resource like which is used in the several organizations can adapt technologies that will help them in the advance thinking or support in the better decision making like in current generation we are able to see the growth in the artificial intelligence,Machine learning,Cyber Security,Data Science etc….Since the HR has the responsibility to build the Organization or a community to develop it and get their shares into the profit.HR has the goal to analyse and predict the data from the subjective aspects but it should be in the Data Analysis.The goal is to analyse work with the objective factors which are attrition the employee to identify the main cause of the problem and contribute the workers decision to the company.This leads to a high cost in terms of productivity, time and money for the company as they were required to hire, rehire, and retrain the new employees to accustom themselves with their new work environment as well as the tasks assigned.

Keywords Machine Learning; Employee Attrition; Prediction Model,Data Analytic,Data Science,Period Prediction,HR Analytic skills,Artificial Intelligence,Neural Networks,Reinforcement Learning,HRIS

1 Introduction

Machine Learning is one of the part in Artificial Intelligence technologies that provide the ability to learn and improve the systems ability to automatically learn and improve from the Neurology or like Human like Intelligence without the Explicit Programming.In the other words the Machine Learning means the developing the computer programs that can access data and use it to learn for themselves **[1]-[4]**.Machine Learning(ML) is currently the fast growing fields of research and has been developed and applied successfully to a wide range of the current generation **[5]-[9]**.This is the study of the real-world domains presents a comparative Analysis of three machine learning algorithms DT,Support Vector Machines(SVM),and Artificial Neural Networks(ANN), to predict the data.

Employee turnover occurs in every company, no matter what their business, whether large-scale or small-scale.Employee turnover or employee churn is a costly problem for companies. The actual cost of replacing an employee is often reasonably large, depending on the experience and skill-sets the employee possesses **[10]**.

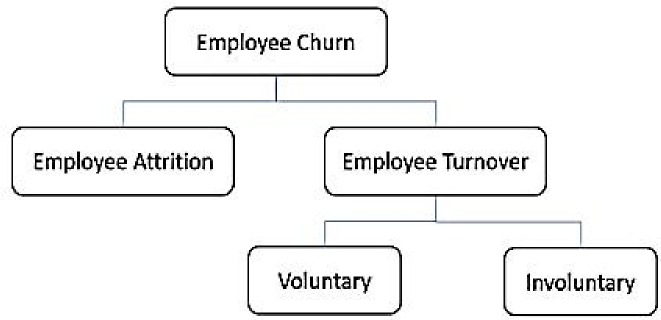
Employee in an organization can mean the reduction of employees through normal means, such as retirement and resignation, clients due to old age, or retrenching them due to change in the target demographics of the organization. The high rate of employee attrition is a major issue in an organization as it greatly impacts them **[11].**The high rate of employee attrition is a major issue in an organization as it greatly impacts them if higher executives or highest-paid employees are to be replaced **[12].**

B Sai Ashish
aldbandhakavi11@gmail.com
Computer Science & Engineering Department,
LPU,Jalandar,Punjab 144411.

This is done by using data from the HRIS of a global retailer and treating the attrition problem as a classification task and modeling it using supervised techniques. The conclusion is reached by contrasting the superior accuracy of the Boost classifier against other techniques and explaining the reason for its superior performance**[13]**.The purpose of this study is to conduct a comparative study to develop machine learning models, i.e., DT, SVM, and ANN, for predicting probable employee attrition and compare between the algorithms in terms of their accuracy and Efficiencies**[14].**However, in literature, much emphasis is laid on prediction using classifier methods with little or no consideration for visualization techniques for data exploration.**[15].**

2 Related work

Human resources are considered an important aspect of an organization, and voluntary employee attrition has been identified as a key issue.In his study focused on identifying employee-related attributes to predict employee attrition using decision tree algorithms. intellectual capital from the employing organization . Most of the literature around turnover categorizes turnover as either voluntary or involuntary. arming techniques to predict turnover thus giving them the vision to take necessary action. Table 1 below briefly documents the literature review findings. Subsequent sections of the paper will highlight the inadequacy of the classifiers recommended here in handling noise of the scale in HRIS. High turnover has several detrimental effects on an organization. It is difficult to replace employees who have niche skill sets or are business domain experts. It affects ongoing work and productivity of existing employees. Acquiring new employees as replacement has its own costs like hiring costs, training costs etc.important performance indicators for data mining algorithms are the accuracy of a classification and the time taken for training. These indicators are mainly useful for selecting the best algorithms for classification or prediction tasks in data mining.The performance measurements observed in many literature reviews are mainly related to finding the best accuracy and speed to build a machine learning model. Briefly documents the literature review findings related to a comparative study on employee attrition using the machine learning classification algorithms.



This presented in their paper an employee churn prediction model by employing five classifiers namely DT, Naive Bayes, ANN and SVM. A subset of employees and customers involved in a specific client unit within a large organization dataset was collected over a year and a half. A random split of 80/20, train and test set were performed on the dataset. Each of the classifiers was set up with unique tuning parameters. Total prediction accuracy was underscored for the five classifiers with SVM having overall performance accuracy of 99.83%.

3 PREDICTION LEARNING

Contrastive learning is a powerful technique in machine learning, particularly in the field of computer vision and natural language processing, but its application in traditional HR data analysis for predicting employee turnover is relatively novel. However, if you intend to explore contrastive learning in this context, here are some unique points to consider in the study methodology:

1. **Representation Learning**:
   - Utilize contrastive learning to learn meaningful representations of employee data by contrasting positive samples (e.g., employees who left the company) with negative samples (e.g., employees who stayed).

2. **Embedding Space**:
   - Explore the creation of an embedding space where similar employees are clustered together based on learned representations, potentially revealing underlying patterns related to turnover.

3. **Unsupervised Learning**:
   - Leverage unsupervised contrastive learning to

discover latent features and relationships within the HR data without relying on explicit labels for employee turnover.

4. **Transfer Learning**:
   - Investigate the potential for transfer learning by pre-training a contrastive model on a related task and fine-tuning it for predicting employee turnover.

5. **Evaluation Metrics**:
   - Define novel evaluation metrics that capture the quality of the learned representations in the context of predicting turnover, potentially considering the separation of turnover and non-turnover instances in the learned embedding space.

6. **Interpretability**:
   - Explore methods to interpret the learned representations and understand the factors that contribute to the contrastive model's predictions regarding employee turnover.

7. **Temporal Aspects**:
   - Account for temporal dynamics in the data and investigate how contrastive learning can capture and leverage temporal patterns related to employee turnover.

8. **Ethical Considerations**:
   - Ensure that the study adheres to ethical guidelines, especially when using employee data for representation learning, and consider the potential impact on individuals.

By considering these unique points in the application of contrastive learning to predict employee turnover using HR data, you can potentially uncover novel insights and patterns that traditional predictive modeling approaches may not capture. However, it's important to note that the application of contrastive learning in this context may require careful experimentation and adaptation of the technique to suit the specific characteristics of HR data and the turnover prediction task.

4 METHODOLOGY

When conducting a study on predicting employee turnover using HR data, the methodology typically involves several key steps and considerations. Here are the unique points to consider in the methodology:

1. **Data Collection**:
   - Gather relevant HR data, including employee demographics, performance evaluations, compensation details, work history, and any other factors that may influence turnover.

2. **Data Preprocessing**:
   - Clean the data to handle missing values, outliers, and inconsistencies. Perform feature engineering to create new variables and transform the data into a suitable format for analysis.

3. **Exploratory Data Analysis (EDA)**:
   - Conduct EDA to gain insights into the distribution of variables, identify patterns, and understand the relationships between different features and the target variable (employee turnover).

4. **Feature Selection**:
   - Use techniques such as correlation analysis, feature importance from models, or domain knowledge to select the most relevant features for predicting turnover.

5. **Model Selection**:
   - Choose appropriate predictive models such as logistic regression, decision trees, random forests, support vector machines, or neural networks based on the nature of the data and the prediction task.

6. **Evaluation Metrics**:
   - Select evaluation metrics such as accuracy, precision, recall, F1 score, and area under the receiver operating characteristic curve (AUC-ROC) to assess the performance of the predictive models.

7. **Handling Imbalance**:
   - Address class imbalance if present in the dataset by using techniques such as oversampling, undersampling, or employing algorithms that inherently handle imbalanced data.

8. **Temporal Considerations**:
   - Account for temporal aspects of the data, such as changes in turnover patterns over time, to ensure that the predictive model captures the evolving dynamics of employee turnover.

9. **Validation Strategy**:
   - Employ cross-validation techniques to assess the generalization ability of the predictive models and avoid overfitting.

10. **Interpretability**:
   - Consider the interpretability of the models to understand the factors influencing turnover and provide actionable insights for HR decision-making.

11. **Ethical Considerations**:
   - Ensure that the study complies with ethical guidelines, especially regarding the use of sensitive employee data and the potential impact on individuals.

By addressing these unique points in the methodology for predicting employee turnover using HR data, you can develop a robust and insightful predictive model that effectively captures the complexities of turnover prediction.

5 Results & discussion

5.1 Dataset

:

1. **IBM HR Analytics Employee Attrition & Performance**: This dataset contains various HR-related attributes such as job satisfaction, work-life balance, performance ratings, and attrition status.

2. **UCI Machine Learning Repository - Employee Attrition Dataset**: This dataset includes features such as job involvement, satisfaction, and performance ratings, along with the target variable indicating whether an employee left the company.

3. **Kaggle - Human Resources Analytics**: This dataset provides information on employee satisfaction, evaluation scores, time spent at the company, and whether the employee left the organization.

4. **HR Analytics: Employee Turnover**: This dataset includes a range of HR metrics such as satisfaction level, number of projects, average monthly hours, and the turnover status of employees.

When using any dataset for research or analysis, it's important to review the terms of use, understand the data's limitations, and ensure compliance with applicable data privacy regulations. If you have a specific dataset in mind or need assistance with data preparation, feel free to provide more details.

Age	Monthly income
Attrition	Monthly rate
Business travel	Number of previous employers
Daily rate	Over 18
Department	Overtime
Distance from home	Per cent salary hike
Education	Performance rating
Education field	Relations satisfaction

5.2 Experimental setup

When setting up an experiment to predict employee turnover using HR data, it's essential to carefully design the methodology to ensure the validity and reliability of the results. Here are some key considerations for the experimental setup:

1. **Data Collection and Preprocessing**:
  - Gather relevant HR data, ensuring that it includes a diverse set of features such as demographics, job satisfaction, performance evaluations, work history, and any other factors that may influence turnover. Clean the data to handle missing values, outliers, and inconsistencies.

2. **Feature Selection and Engineering**:
  - Conduct exploratory data analysis (EDA) to identify relevant features and relationships. Engineer new features if necessary, such as tenure, performance trends, or satisfaction scores.

3. **Experimental Groups**:
  - Define the experimental groups, such as employees who left the company (turnover) and those who stayed, ensuring a balanced representation of both groups in the dataset.

4. **Temporal Considerations**:

- Account for temporal aspects, such as changes in turnover patterns over time, and consider the impact of historical trends on the predictive models.

	Attrition	Age	BusinessTravel	DailyRate	Department	DistanceFromHome	Education	EducationField	EnvironmentSatisfaction	Gender	JobInvolvem
count	1470	1470.00	1470	1470.00	1470	1470.00	1470.00	1470	1470.00	1470	1470
unique	2	NaN	3	NaN	3	NaN	NaN	6	NaN	2	N
top	No	NaN	Travel_Rarely	NaN	Research & Development	NaN	NaN	Life Sciences	NaN	Male	N
freq	1233	NaN	1043	NaN	961	NaN	NaN	606	NaN	882	N
mean	NaN	36.92	NaN	802.49	NaN	9.19	2.91	NaN	2.72	NaN	2
std	NaN	9.14	NaN	403.51	NaN	8.11	1.02	NaN	1.09	NaN	0
min	NaN	18.00	NaN	102.00	NaN	1.00	1.00	NaN	1.00	NaN	1
25%	NaN	30.00	NaN	465.00	NaN	2.00	2.00	NaN	2.00	NaN	2
50%	NaN	36.00	NaN	802.00	NaN	7.00	3.00	NaN	3.00	NaN	3
75%	NaN	43.00	NaN	1157.00	NaN	14.00	4.00	NaN	4.00	NaN	3
max	NaN	60.00	NaN	1499.00	NaN	29.00	5.00	NaN	4.00	NaN	4

5. **Model Selection**:
 - Choose appropriate predictive models, considering techniques such as logistic regression, decision trees, random forests, support vector machines, or neural networks based on the nature of the data and the prediction task.

6. **Training and Testing Split**:
 - Split the dataset into training and testing sets, ensuring that the testing set is representative of the population and contains a sufficient number of turnover instances for reliable evaluation.

7. **Cross-Validation**:
 - Employ cross-validation techniques to assess the generalization ability of the predictive models and avoid overfitting.

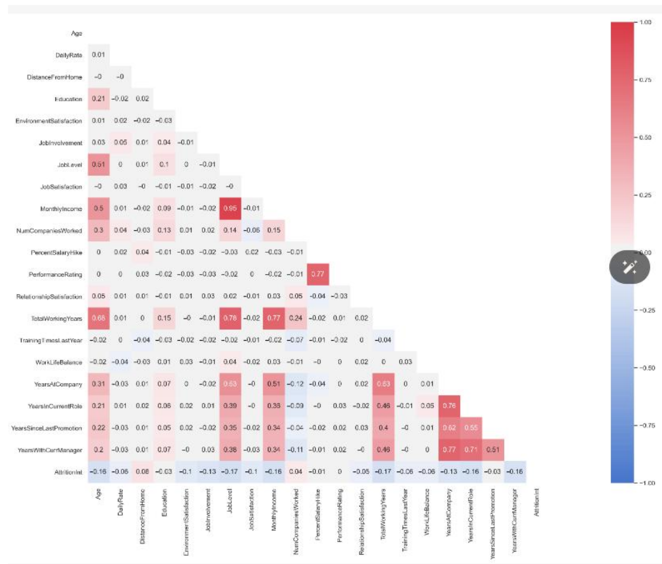
8. **Evaluation Metrics**:
 - Select appropriate evaluation metrics such as accuracy, precision, recall, F1 score, and area under the receiver operating characteristic curve (AUC-ROC) to assess the performance of the predictive models.

9. **Ethical Considerations**:
 - Ensure that the study complies with ethical guidelines, especially regarding the use of sensitive employee data and the potential impact on individuals.

10. **Experimental Controls**:
 - Consider the inclusion of control variables or groups to account for potential confounding factors that may influence turnover predictions, such as external market conditions or organizational changes.

5.3 Quantitative results

1. **Model Performance Metrics**:
 - Accuracy: The proportion of correctly predicted turnover and non-turnover instances.
 - Precision: The ratio of true positive predictions to the total number of positive predictions, indicating the model's ability to avoid false positives.
 - Recall: The ratio of true positive predictions to the total number of actual positive instances, demonstrating the model's ability to capture turnover cases.
 - F1 Score: The harmonic mean of precision and recall, providing a balanced measure of the model's performance.
 - Area Under the Receiver Operating Characteristic Curve (AUC-ROC): A measure of the model's ability to distinguish between turnover and non-turnover instances across different thresholds.



2. **Feature Importance**:
 - Quantify the importance of different features in predicting turnover using methods such as coefficients from logistic regression, feature importances from tree-based models, or SHAP (SHapley Additive exPlanations) values.

3. **Confusion Matrix**:
 - Present the confusion matrix to illustrate the model's performance in terms of true positives, true negatives, false positives, and false negatives.

4. **Cross-Validation Results**:
 - Include cross-validation results to demonstrate the consistency of the model's performance across

different folds of the dataset.

5. **Comparative Analysis**:
- Compare the performance of different predictive models if multiple models were evaluated, highlighting the strengths and weaknesses of each approach.

6. **Temporal Analysis**:
- If applicable, present quantitative results related to temporal patterns in turnover prediction, such as changes in model performance over time or the impact of historical trends.

7. **Statistical Significance**:
- If relevant, include statistical tests to assess the significance of differences in model performance or feature importance.

8. **Scalability and Efficiency**:
- If relevant, provide quantitative insights into the scalability and computational efficiency of the predictive models, especially in large-scale HR data settings.

5.4 Qualitative results

1. **Feature Analysis**:
- Qualitatively describe the most influential features in predicting turnover, highlighting their potential impact on employee retention and satisfaction.

2. **Case Studies**:
- Provide qualitative case studies or narratives of individual employees or groups of employees to illustrate the factors that led to turnover, potentially highlighting common themes or patterns.

3. **Employee Segmentation**:
- Describe qualitative insights gained from segmenting employees based on turnover predictions, such as identifying groups with high turnover risk and understanding the unique characteristics of each segment.

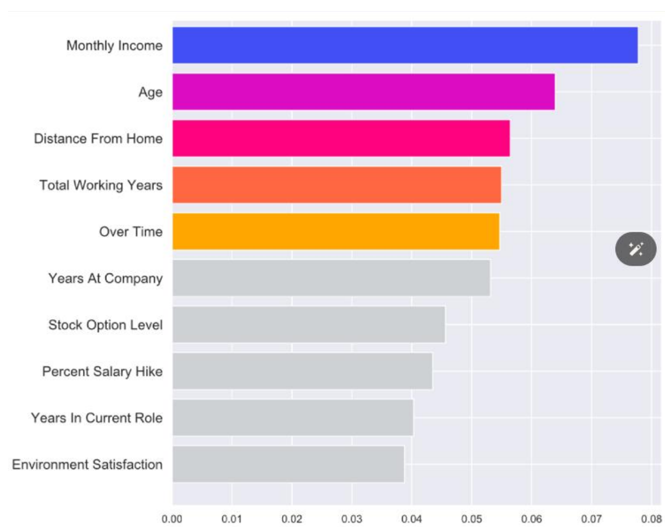
4. **Interpretation of Feature Importance**:
- Qualitatively interpret the importance of different features in predicting turnover, providing insights into the underlying reasons for their impact on employee retention.

5. **Model Limitations and Assumptions**:
- Discuss the qualitative limitations and assumptions of the predictive models, including potential biases, contextual factors, and areas where the models may not fully capture the complexities of turnover prediction.

6. **Temporal Patterns and Trends**:
- Qualitatively analyze temporal patterns in turnover prediction, such as changes in turnover rates over time or the impact of external events on employee retention.

7. **Employee Feedback and Surveys**:
- If available, incorporate qualitative feedback from employees or surveys related to job satisfaction, work environment, and reasons for considering turnover.

8. **Organizational Insights**:
- Provide qualitative insights into the implications of turnover predictions for organizational decision-making, potentially including recommendations for interventions or policies to address turnover risk factors.



By presenting these qualitative results in a descriptive and insightful manner, you can provide a comprehensive understanding of the factors influencing employee turnover and the implications for HR management and organizational strategies.

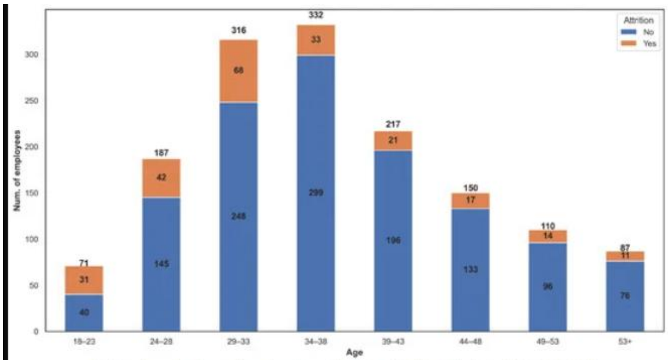
Collect the employee dataset, which consists of current and past employee observations (Section 3.2.1);

Apply various data cleaning techniques to prepare the dataset (Section 3.2.2);

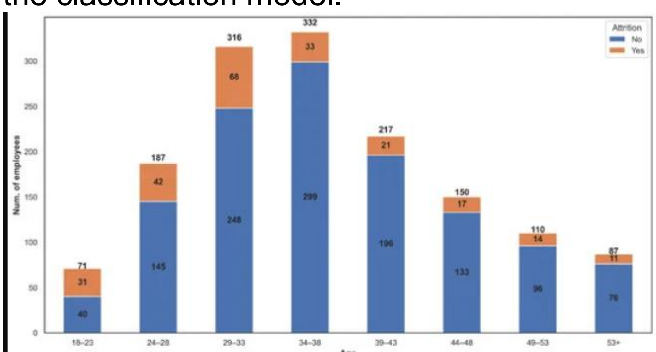
Start a descriptive analysis of data to detect the key factors and trends that contribute to attrition (Section 3.3);

Elaborate the dataset for the training and testing phase and try several classification algorithms to process it (Section 4);

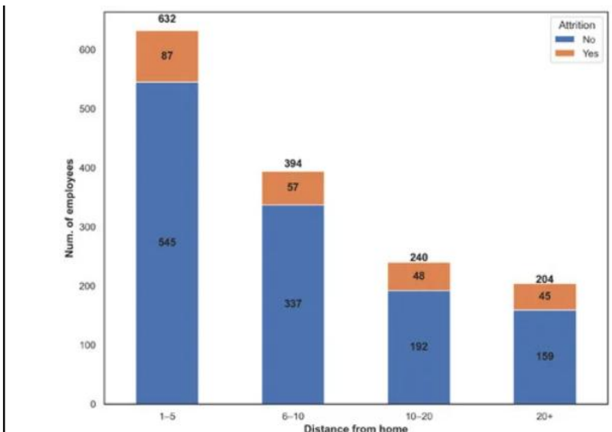
Based on the results collected with test data, compare many performance metrics of machine learning models and select which model best fits and gives the most accurate results for the given problem (Section 5) and release HR support software that implements the classification model.



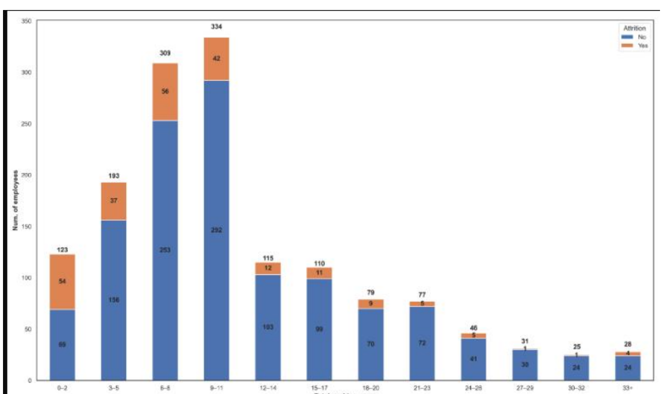
	AgeBins	Num. of employees	Num. of employees with attrition	% attrition in the cluster	% of the total attrition
0	18–23	71	31	43.7	13.1
1	24–28	187	42	22.5	17.7
2	29–33	316	68	21.5	28.7
3	34–38	332	33	9.9	13.9
4	39–43	217	21	9.7	8.9
5	44–48	150	17	11.3	7.2
6	49–53	110	14	12.7	5.9
7	53+	87	11	12.6	4.6



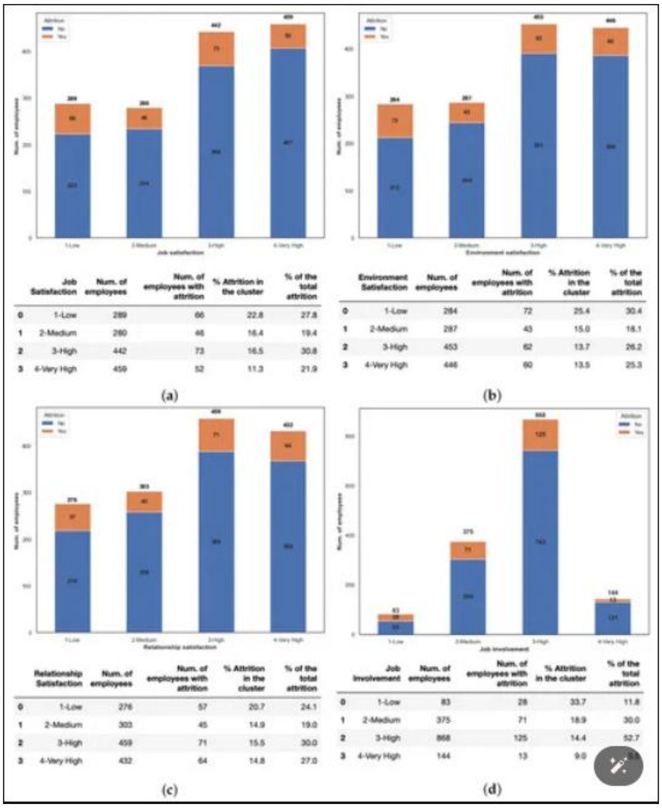
	AgeBins	Num. of employees	Num. of employees with attrition	% attrition in the cluster	% of the total attrition
0	18–23	71	31	43.7	13.1
1	24–28	187	42	22.5	17.7
2	29–33	316	68	21.5	28.7
3	34–38	332	33	9.9	13.9
4	39–43	217	21	9.7	8.9
5	44–48	150	17	11.3	7.2
6	49–53	110	14	12.7	5.9
7	53+	87	11	12.6	4.6



	DistanceFromHomeBins	Num. of employees	Num. of employees with attrition	% attrition in the cluster	% of the total attrition
0	1–5	632	87	13.8	36.7
1	6–10	394	57	14.5	24.1
2	10–20	240	48	20.0	20.3
3	20+	204	45	22.1	19.0



	TotalWorkingYearsBins	Num. of employees	Num. of employees with attrition	% attrition in the cluster	% of the total attrition
0	0–2	123	54	43.9	22.8
1	3–5	193	37	19.2	15.6
2	6–8	309	56	18.1	23.6
3	9–11	334	42	12.6	17.7
4	12–14	115	12	10.4	5.1
5	15–17	110	11	10.0	4.6
6	18–20	79	9	11.4	3.8
7	21–23	77	5	6.5	2.1
8	24–26	46	5	10.9	2.1
9	27–29	31	1	3.2	0.4
10	30–32	25	1	4.0	0.4
11	33+	28	4	14.3	1.7

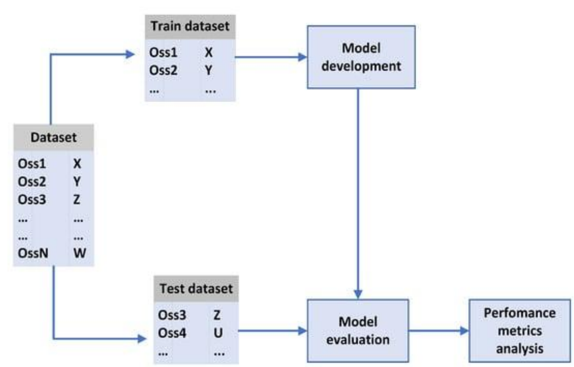


After identifying the objectives and adequately preparing and analysing the dataset to be used, we proceeded with the design of the prediction model to identify employees that would potentially leave the company. In the construction phase of a model that implements a supervised learning algorithm, it was necessary to have a training-set available that consisted of instances of an already classified population (target), in order to train the model to classify new observations, which will constitute the test-set (in which the attribute representing the class was missing). Then, the model must be trained on a consistent number of observations in order to refine its prediction ability. The precision of the machine learning algorithms increases with the amount of data available during training. Ideally, one would have two distinct datasets: one for training and a second to be used as a test. As two dedicated datasets were not available in this case, the original dataset was divided into two parts with a 70:30 ratio, one used for training and one used for testing (see **Figure 12**):

6. Model Building

The modelling process consists in selecting models that are based on various machine learning techniques used in the experimentation. In this case various predictive models were used such as those based on decision tree, Bayesian method, logistic regression and SVM. The goal is to identify the best classifier for the analysed problem. Each classifier must therefore be trained on the featured set and the classifier with the best classification results is used for prediction. The classification algorithms taken into consideration are:

- Gaussian Naive Bayes,
- Naive Bayes classifier for multivariate Bernoulli models,
- Logistic Regression classifier,
- K-nearest neighbours (K-NN),
- Decision tree classifier,
- Random forest classifier,
- Support Vector Machines (SVM) classification,
- Linear Support Vector Machines (LSVM) classification.



Train set contained 70% of the dataset. This information was dedicated to the training phase in order to allow the model to learn the relationships hidden in the data; the train-set contains 1029 observations;

Test set contained the remaining 30%. This information was dedicated to the test and validation phase in order to evaluate the general performance of the model and to calculate errors between predicted and actual results; the test-set contains 441 observations.

In addition, the newly created train and test datasets were further divided to extract the target variable (“Attrition”); the label was stored in a dedicated dataset (y) separating it from the dataset (X) containing the rest of the variables:

X, containing all independent variables;

y, containing the dependent variable, i.e., “Attrition”;

When evaluating the performance of a model, it is important to perform independent evaluation tests and to use multiple observations in assessment in order to obtain more reliable and accurate indicators of errors. Therefore, we adopted the following two techniques for a better error estimation:

Holdout: When the datasets are in the split phase, it is essential to keep the same distribution of target variables within both the training and test datasets. Thus, it is necessary to avoid that a random subdivision can alter the proportion of the classes present in the training and test datasets from that in the original. The target (“attrition” attribute) is a binary variable with 84% “No” and 16% “Yes”, both datasets kept the same proportion after the split.

Cross-validation: We adopted this technique to prevent over-fitting problems and to simplify the model. The training-set was randomly divided into five parts (k)—one was used as a validation-set and the other k-1s as training-sets, repeating the procedure k times. In each of the iterations, a different part was taken as the validation-set and finally the average prediction error was obtained by assessing the average errors in the k iterations performed on each k-validation set (see **Figure 13**).

Train set contained 70% of the dataset. This information was dedicated to the training phase in order to allow the model to learn the relationships hidden in the data; the train-set contains 1029 observations;

Test set contained the remaining 30%. This information was dedicated to the test and validation phase in order to evaluate the general performance of the model and to calculate errors between predicted and actual results; the test-set contains 441 observations.

In addition, the newly created train and test datasets were further divided to extract the target variable (“Attrition”); the label was stored in a dedicated dataset (y) separating it from the dataset (X) containing the rest of the variables:

X, containing all independent variables;

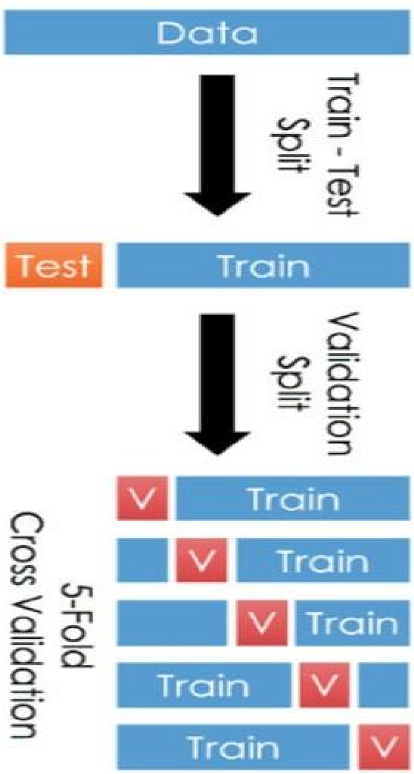
y, containing the dependent variable, i.e., “Attrition”;

When evaluating the performance of a model, it is important to perform independent evaluation tests and to use multiple observations in assessment in order to obtain more reliable and accurate indicators of errors. Therefore, we adopted the following two techniques for a better error estimation:

Holdout: When the datasets are in the split phase, it is essential to keep the same distribution of target variables within both the training and test datasets. Thus, it is necessary to avoid that a random subdivision can alter the proportion of the classes present in the training and test datasets from that in the original. The target (“attrition” attribute) is a binary variable with 84% “No” and 16% “Yes”, both datasets kept the same proportion after the split.

Cross-validation: We adopted this technique to prevent over-fitting problems and to simplify the model. The training-set was randomly divided into five parts (k)—one was used as a validation-set and the other k-1s as training-sets, repeating the procedure k times. In each of the iterations, a different part was taken as the validation-set and finally the average prediction error was obtained by assessing the average errors in the k iterations performed on each k-validation set (see **Figure 13**).

$$Z_i = \frac{(x_i - \mu)}{\sigma}$$



7. Conclusion

This work tried to provide answers to some of the common questions of responsible human resources management:

What are the key indicators that signal that an employee will leave the company?

What is the probability that an employee will leave the company?

To this aim, we applied some machine learning techniques in order to identify the factors that may contribute to an employee leaving the company and, above all, to predict the likelihood of individual employees leaving the company. First, we assess statistically the data and then we classified them. The dataset was processed, dividing it into the training phase and the test phase, guaranteeing the same distribution of the target variable (through the holdout technique). We selected various classification algorithms and, for each of them, we carried out the training and validation phases. To evaluate the algorithm's performance, the predicted results were collected and fed into the respective confusion matrices. From these it was possible to calculate the basic metrics necessary for an overall evaluation (precision, recall, accuracy, f1 score, ROC curve, AUC, etc.) and to identify the most suitable classifier to predict whether an employee was likely to leave the company. The algorithm that produced the best results for the available dataset was the Gaussian Naïve Bayes classifier: it revealed the best recall rate (0.540.54), a metric that measures the ability of a classifier to find all the positive instances, and achieved an overall false negative rate equal to 4.5%4.5% of the total observations. Results obtained by the proposed automatic predictor demonstrate that the main attrition variables are monthly income, age, overtime, distance from home. The results obtained from the data analysis represent a starting point in the development of increasingly efficient employee attrition classifiers. The use of more numerous datasets or simply to update it periodically, the application of feature engineering to identify new significant characteristics from the dataset and the availability of additional information on employees would improve the overall knowledge of the reasons why employees leave their companies and, consequently, increase the time available to personnel departments to assess and plan the tasks required to mitigate this risk (e.g., retention activities, employee substitution and/or task redistribution).

Other existing economic evidence underlines the role played by outside opportunities on the labour market in employees utility in the current job and turnover intention [19,34,35,36]. In future research it is possible to improve the analysis by considering new employees' opportunities as well as adverse working conditions (e.g., harm and hazard) and poor promotion prospects, discrimination and low social support, that are positively related to employees' turnover intention [37,38].

8. References

- Cockburn, I.; Henderson, R.; Stern, S. The Impact of Artificial Intelligence on Innovation. In *The Economics of Artificial Intelligence: An Agenda*; University of Chicago Press: Chicago, IL, USA, 2019; pp. 115–146. [Google Scholar]
- Jarrahi, M. Artificial intelligence and the future of work: Human-AI symbiosis in organizational decision making. *Bus. Horiz.* **2018**, *61*, 577–586. [Google Scholar] [CrossRef]
- Yanqing, D.; Edwards, J.; Dwivedi, Y. Artificial intelligence for decision making in the era of Big Data. *Int. J. Inf. Manag.* **2019**, *48*, 63–71. [Google Scholar]
- Paschek, D.; Luminosu, C.; Dra, A. Automated business process management-in times of digital transformation using machine learning or artificial intelligence. In *MATEC Web of Conferences*; EDP Sciences: Les Ulis, France, 2017; Volume 121. [Google Scholar]
- Varian, H. *Artificial Intelligence, Economics, and Industrial Organization*; National Bureau of Economic Research: Cambridge, MA, USA, 2018. [Google Scholar]
- Vardarlier, P.; Zafer, C. Use of Artificial Intelligence as Business Strategy in Recruitment Process and Social Perspective. In *Digital Business Strategies in Blockchain Ecosystems*; Springer: Berlin/Heidelberg, Germany, 2019; pp. 355–373. [Google Scholar]
- Gupta, P.; Fernandes, S.; Manish, J. Automation in Recruitment: A New Frontier. *J. Inf. Technol. Teach. Cases* **2018**, *8*, 118–125. [Google Scholar] [CrossRef]
- Geetha, R.; Bhanu Sree Reddy, D. Recruitment through artificial intelligence: A conceptual study. *Int. J. Mech. Eng. Technol.* **2018**, *9*, 63–70. [Google Scholar]
- Syam, N.; Sharma, A. Waiting for a sales renaissance in the fourth industrial revolution: Machine learning and artificial intelligence in sales research and practice. *Ind. Mark. Manag.* **2018**, *69*, 135–146. [Google Scholar] [CrossRef]
- Mishra, S.; Lama, D.; Pal, Y. Human Resource Predictive Analytics (HRPA) For HR Management in Organizations. *Int. J. Sci. Technol. Res.* **2016**, *5*, 33–35. [Google Scholar]
- Jain, N.; Maitri. Big Data and Predictive Analytics: A Facilitator for Talent Management. In *Data Science Landscape*; Springer: Singapore, 2018; pp. 199–204. [Google Scholar]
- Boushey, H.; Glynn, S.J. There Are Significant Business Costs to Replacing Employees. *Cent. Am. Prog.* **2012**, *16*, 1–9. [Google Scholar]
- Martin, L. How to retain motivated employees in their jobs? *Econ. Ind. Democr.* **2018**, *34*, 25–41. [Google Scholar] [CrossRef]
- Wood, S.; Van Veldhoven, M.; Croon, M.; de Menezes, L.M. Enriched job design, high involvement management and organizational performance: The mediating roles of job satisfaction and wellbeing. *Hum. Relat.* **2012**, *65*, 419–446. [Google Scholar] [CrossRef]
- Zelenski, J.M.; Murphy, S.A.; Jenkins, D.A. The happy-productive worker thesis revisited. *J. Happiness Stud.* **2008**, *9*, 521–537. [Google Scholar] [CrossRef]
- Clark, A.E. What really matters in a job? Hedonic measurement using quit data. *Labour Econ.* **2001**, *8*, 223–242. [Google Scholar] [CrossRef]
- Clark, A.E.; Georgellis, Y.; Sanfey, P. Job satisfaction, wage changes, and quits: Evidence from Germany. *Res. Labor Econ.* **1998**, *17*, 95–121. [Google Scholar]
- Delfgaauw, J. The effect of job satisfaction on job search: Not just whether, but also where. *Labour Econ.* **2007**, *14*, 299–317. [Google Scholar] [CrossRef] [Green Version]
- Green, F. Well-being, job satisfaction and labour mobility. *Labour Econ.* **2010**, *17*, 897–903. [Google Scholar] [CrossRef] [Green Version]
- Kristensen, N.; Westergaard-Nielsen, N. Job satisfaction and quits – which job characteristics matters most? *Dan. Econ. J.* **2006**, *144*, 230–249. [Google Scholar]
- Marchington, M.; Wilkinson, A.; Donnelly, R.; Kynighou, A. *Human Resource Management at Work*; Kogan Page Publishers: London, UK, 2016. [Google Scholar]
- Van Reenen, J. Human resource management and productivity. In *Handbook of Labor Economics*; Elsevier: Amsterdam, The Netherlands, 2011. [Google Scholar]
- Deepak, K.D.; Guthrie, J.; Wright, P. Human Resource Management and Labor Productivity: Does Industry Matter? *Acad. Manag. J.* **2005**, *48*, 135–145. [Google Scholar]
- Gordini, N.; Veglio, V. Customers churn prediction and marketing retention strategies. An application of support vector machines based on the AUC parameter-selection technique in B2B e-commerce industry. *Ind. Mark. Manag.* **2016**, *62*, 100–107. [Google Scholar] [CrossRef]
- Keramati, A.; Jafari-Marandi, R.; Aliannejadi, M.; Ahmadian, I.; Mozaffari, M.; Abbasi, U. Improved churn prediction in telecommunication industry using data mining techniques. *Appl. Soft Comput.* **2014**, *24*, 994–1012. [Google Scholar] [CrossRef]
- Alao, D.; Adeyemo, A. Analyzing employee attrition using decision tree algorithms. *Comput. Inf. Syst. Dev. Inf. Allied Res. J.* **2013**, *4*, 17–28. [Google Scholar]
- Nagadevara, V. Early Prediction of Employee Attrition in Software Companies-Application of Data Mining Techniques. *Res. Pract. Hum. Resour. Manag.* **2008**, *16*, 2020–2032. [Google Scholar]
- Rombaut, E.; Guerry, M.A. Predicting voluntary turnover through Human Resources database analysis. *Manag. Res. Rev.* **2018**, *41*, 96–112. [Google Scholar] [CrossRef]

- Usha, P.; Balaji, N. Analysing Employee attrition using machine learning. *Karpagam J. Comput. Sci.* **2019**, *13*, 277–282. [**Google Scholar**]
- Ponnuru, S.; Merugumala, G.; Padigala, S.; Vanga, R.; Kantapalli, B. Employee Attrition Prediction using Logistic Regression. *Int. J. Res. Appl. Sci. Eng. Technol.* **2020**, *8*, 2871–2875. [**Google Scholar**] [**CrossRef**]
- Microsoft Docs: Team Data Science Process. Available online: **https://docs.microsoft.com/it-it/azure/machine-learning/team-data-science-process/** (accessed on 31 October 2020).
- IBM HR Analytics Employee. Available online: **https://www.kaggle.com/pavansubhasht/ibm-hr-analytics-attrition-dataset** (accessed on 31 October 2020).
- CrowdFlower. Data Science Report. 2016. Available online: **https://visit.figure-eight.com/data-science-report.html** (accessed on 31 October 2020).
- Antecol, H.; Cobb-Clark, D. Racial harassment, job satisfaction, and intentions to remain in the military. *J. Popul. Econ.* **2009**, *22*, 713–738. [**Google Scholar**] [**CrossRef**] [**Green Version**]
- Böckerman, P.; Ilmakunnas, P. Job disamenities, job satisfaction, quit intentions, and actual separations: Putting the pieces together. *Ind. Relations* **2009**, *48*, 73–96. [**Google Scholar**] [**CrossRef**] [**Green Version**]
- Theodossiou, I.; Zangelidis, A. Should I stay or should I go? The effect of gender, education and unemployment on labour market transitions. *Labour Econ.* **2009**, *16*, 566–577. [**Google Scholar**] [**CrossRef**]
- Böckerman, P.; Ilmakunnas, P.; Jokisaari, M.; Vuori, J. Who stays unwillingly in a job? A study based on a representative random sample of employees. *Econ. Ind. Democr.* **2013**, *34*, 25–41. [**Google Scholar**] [**CrossRef**] [**Green Version**]
- Griffeth, R.W.; Hom, P.W.; Gaertner, S. A meta-analysis of antecedents and correlates of employee turnover: Update, moderator tests, and research implications for the next millennium. *J. Manag.* **2000**, *26*, 463–488. [**Google Scholar**] [**CrossRef**]