



EMPLOYEE ATTRITION PREDICTION

Mr. N. Arikaran[1], David jackson J[2], Janakan H.M[3], Antony jerome[4]

¹ Assistant Professor, Department of Computer Science & Engineering,
Manakula Vinayagar Institute of Technology, Puducherry, India.

^{2,3,4} UG Scholar, Department of Computer Science & Engineering,

^{2,3,4} Manakula Vinayagar Institute of Technology, Puducherry, India.

Abstract: The goal of this project is to use the k-Nearest Neighbor algorithm to forecast whether or not a company's employee will leave. We use, among other things, the number of years an employee has worked for the company as well as the appraisal of employee performance on a monthly basis. Logistic regression, decision trees, and artificial neural networks (ANNs) are further methods for solving this issue. The dataset was divided, with 70% of it being used to train the algorithm and 30% of it being used to test it, yielding an accuracy of 94.32%.

Employee turnover is a prevalent issue faced by organizations, where experienced and valuable employees consistently choose to leave. This study aims to develop a model that can predict whether an employee will leave the organization. By leveraging a substantial amount of employee-related data, we propose a machine learning-based approach to efficiently address this attrition problem. Specifically, we employed four algorithms, namely Decision Tree, K-Nearest Neighbor, Support Vector Machine, and Light Gradient Boosting Machine, to build our predictive model. The accuracy of these algorithms was evaluated using the IBM dataset for both training and testing. By implementing this system, organizations can enhance their human resource management practices and reduce costs. This project provides valuable insights that can aid in predicting employee attrition.

1.INTRODUCTION

Advent of globalization has compelled businesses and their leaders to adopt a global mindset and approach to gain a competitive advantage. In the global landscape, the success of competitive organizations hinges on the uniqueness of their human resources and effective human resource management systems. Just as an engine propels an automobile towards its destination, a firm's resources propel it towards its goals. People form the core of every organization. Acquiring their services, developing their skills, motivating them to perform at high levels, and ensuring their commitment to the organization are indispensable for achieving organizational objectives. Human Resource Management (HRM) is a management discipline that focuses on overseeing and nurturing the workforce within an organization and managing their relationships. The main objective of HRM is to ensure that the organization has a capable and motivated workforce. Armstrong (2006) defines HRM as a strategic and well-structured

approach to managing the organization's most valuable assets—the individuals employed there—who play a crucial role, both individually and collectively, in achieving the organization's goals. In contrast, personnel management centers on the management of individuals and their relationships within the organization. Human Resource Management (HRM) aims to unite individuals within an enterprise, empowering them to contribute their best towards its success as individuals and as part of a cohesive team. The ultimate objective of HRM is to enable the organization to achieve success through its people. The human capital of an organization comprises the individuals who work there and upon whom the business's success relies. Human capital is often described as the embodiment of the human element within the organization, encompassing the collective intelligence, skills, and expertise that give the organization its unique identity.

This paragraph emphasizes the importance of human capital, referring to the individuals who make up an organization's workforce, as the organization's most valuable asset. It highlights the need for businesses to invest in their human capital to ensure the long-term sustainability and growth of the organization. The paragraph emphasizes that the success of an organization is driven by the individuals comprising its workforce. The collective intelligence, skills, and expertise of these individuals contribute to the organization's unique identity and competitive advantage. Recognizing the significance of human capital, businesses are encouraged to allocate resources towards nurturing and developing their workforce. This involves providing opportunities for learning, skill development, and innovation. The paragraph highlights that the human elements within the organization possess the ability to learn, adapt, and innovate, and by harnessing this potential and motivation, the organization can ensure its long-term survival. To effectively manage and maximize the potential of their human capital, organizations are advised to focus on Human Resource Management (HRM) practices. This includes strategies for attracting, retaining, and motivating a skilled and dedicated workforce that aligns with the organization's goals and objectives. Overall, the paragraph stresses the pivotal role of human capital in driving organizational success and emphasizes the need for businesses to invest in their workforce through HRM practices to secure sustainability, growth, and competitive advantage. This means taking steps to assess and satisfy future people needs and to enhance and develop the inherent capabilities of people – their contributions, potential and employ ability – by providing learning and continuous development opportunities. It also means engaging in talent management – the process of acquiring and nurturing talent. Knowledge management is any process or practice of creating, acquiring, capturing, sharing and using knowledge, wherever it resides, to enhance learning and performance in organizations. The objective of HRM is to facilitate the growth of organization-specific knowledge and skills that arise from the process of organizational learning.

HRM endeavors to boost motivation, job engagement, and commitment by implementing policies and practices that recognize and reward individuals for their accomplishments, contributions, and level of expertise. On the other hand, Machine Learning refers to a computer algorithm system that can autonomously improve and learn from examples without the need for explicit programming by a developer. Machine learning, a subset of artificial intelligence, combines data with statistical tools to generate predictive outputs that lead to actionable

insights. The key breakthrough lies in the concept that machines can independently learn from data examples, resulting in accurate outcomes. Machine learning is closely connected to data mining and Bayesian predictive modeling. By receiving data as input, machines utilize algorithms to formulate answers. One common application of machine learning is providing recommendations, such as the movie or series suggestions on Netflix based on users' viewing history. Tech companies employ unsupervised learning to enhance user experiences through personalized recommendations. Machine learning is also utilized in various tasks like fraud detection, predictive maintenance, portfolio optimization, and task automation. Contrasting traditional programming, where programmers manually code rules based on logical foundations, machine learning automates the learning process. This alleviates the need to explicitly define rules for complex systems, making maintenance more sustainable in the long run. Traditional programming and machine learning exhibit fundamental differences in their approaches. In traditional programming, a programmer is responsible for manually coding all the rules in collaboration with domain experts relevant to the software's industry. Each rule is formulated based on a logical foundation, ensuring that the machine executes an output in accordance with the specified logical statement. However, as the system becomes more intricate, the number of rules that need to be written multiplies significantly. This proliferation of rules can rapidly lead to an unsustainable situation in terms of maintenance. The challenges of traditional programming arise from the need to anticipate and account for every possible scenario and condition explicitly. This process requires the programmer to possess extensive expertise in the domain and painstakingly encode each rule, covering all conceivable situations. As the system grows larger and more complex, the number of rules escalates exponentially, making it increasingly difficult and time-consuming to manage and update the code-base. This approach is not only labor-intensive but also prone to errors and oversights, leading to potential limitations in the system's functionality.

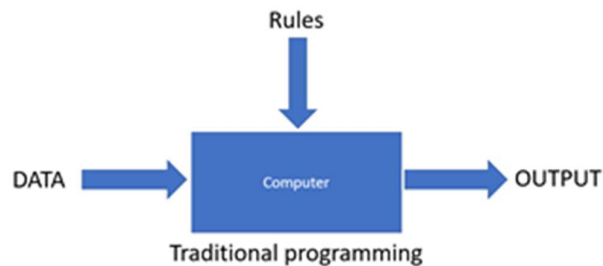


FIG 1: Traditional Programming

Machine learning is supposed to overcome this issue. The machine learns how the input and output data are correlated and it writes a rule. The programmers do not need to write new rules each time there is new data. The algorithms adapt in response to new data and experiences to improve efficacy overtime. Machine Learning Machine learning is the brain where all the learning takes place. The way the machine learns is similar to the human being. Humans learn from experience. The more we know, the more easily we can predict. By analogy, when we face an unknown situation, the likelihood of success is lower than the known situation. Machines are trained the same. To make an accurate prediction, the machine sees an example. When we give the machine a similar example, it can figure out the outcome. However, like a human, if its feed a previously unseen example, the machine has difficulties to predict. The

core objective of machine learning is the learning and inference. The first way the machine learns is by identifying patterns. This discovery is made thanks to the data. One crucial part of the data scientist is to choose carefully which data to provide to the machine. The list of attributes used to solve a problem is called a feature vector. You can think of a feature vector as a subset of data that is used to tackle a problem. The machine uses some fancy algorithms to simplify the reality and transform this discovery into a model. Therefore, the learning stage is used to describe the data and summarize it into a model.

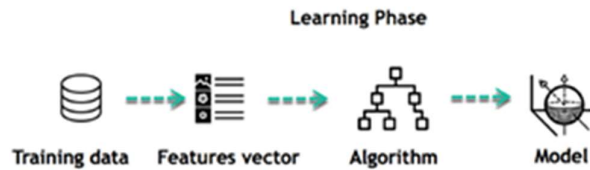


FIG 2: Learning Phase

For instance, the machine is trying to understand the relationship between the wage of an individual and the likelihood to go to a fancy restaurant. It turns out the machine finds a positive relationship between wage and going to a high-end restaurant: This is the model. When the model is built, it is possible to test how powerful it is on never-seen-before data. The new data are transformed into a features vector, go through the model and give a prediction. This is all the beautiful part of machine learning. There is no need to update the rules or train again the model. You can use the model previously trained to make inference on new data.

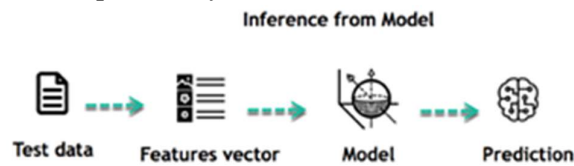


FIG 3: Inference from Model

The price of attrition is printed due to the fact the type of team of workers UN enterprise leaves a agency during a particular time frame divided with the aid of using the everyday general type of team of workers over that very identical time frame. It's overpriced, non-efficient and frustrating. Harvard Business overview calculable the fee of turnover at round companion tiers fiftieth|100 and fiftieth} of an employee's annual salary and could have an impact at the 400th of enterprise profits.

ATTRITION CYCLE



FIG 4: Decreasing cycle of Attrition

Individuals getaway creating others paintings additional diligently
This provides additional whittling down, that provides to elevated prices and decrease-income
This powers additional rate lower and somberness measures
This, thus, makes operative additional difficult, inflicting the standard entertainers with the
most outer events to depart Representative whittling down, the first-rate cause for concern for
corporations, levels among fifteen keep with cent and two hundredths. AN enterprise is not
injured thanks to the very fact, somebody, especially, has left. The enterprise is injured thanks
to the very fact he/she gets eliminate positive data, and there is also no organized substitution
at the lookout. Weakening, thus, is certainly currently not a frightful marvel. it's been
acknowledged to exist from the beginning. Be that as a result of it should conjointly, whereas
occupations are scant, the innovation additional First State become abundant less fast, the
planned consistent loss becomes very little and businesses oversaw it. withal, with innovation
changing quick and labour charges increasing, carrying down is excessive and broken severely.
Huge gamer oftentimes use coins cap potential to draw in cap potential from additional modest
gamer. Organizations likewise create use of the 'place' lure to tug in workers. whereas a selected
degree of labour turnover is appealing to keep up new blood returning in, and eliminating
deadwood, higher prices are positively horrific tips of AN affiliation's manner of lifestyles and
folks rehearses.

It is a check to seek out the simplest cap potential inner limitations, as an example, place,
nature of paintings, pay and edges. New alumni turning into a member of AN affiliation create
it a spotlight getaway within the primary year. They place along with themselves currently not
for gambling out their gift space of employment but as a substitute for drawing in another one.
various an amount, there is also a relocation from additional businesses to additional modest
businesses likewise, primarily given the infamy related to a selected venture or a selected client.
currently and once more, additional modest businesses are shockingly higher paymasters than
larger brands. The success paces of repairs applications are lots higher if the management
makes use of a second, employee-centered, technique and is prepared to form a contribution
belongings for the equivalent. The PRESENT state of affairs Steady loss is motivated with the
help of victimization good financial things Economy blast Positive data about the economic
system-after the four-year worsening, the financial system offers off AN influence of being at
the recuperate and is giving warning signs of attainable pressure. money tips uphold this

recuperation: improvement charge additional than V-E Day

The modern financial powers are organizing weather at the money for ability business enterprise collapse - sudden worker consistent loss and turnover at a degree several businesses have in no manner full-fledged. Organizations at one time debilitated with the help of victimization scaling lower back and cost-reducing should be put in to treat the symptoms and symptoms and create a flow into if they are to endure, well abundant less dominate. Studies have exposed the accompanying:83 keep with cent of representatives are probably reaching to rummage around for new paintings as a result of the financial system blasts. XL viii keep with cent of chiefs are probably reaching to rummage around for new paintings with the enhancing financial system. cardinal keep with cent of those chiefs is with success looking. 56 keep with cent of hour professionals showed, all matters thought of, intentional turnover may ascend thanks to the enhancing financial system. Source: worker Vulnerable Study with the help of victimization NFO Republic of India Albeit a part of these variables is on the far side the enterprise's instant management. At the issue, whereas top-notch cap potential strolls, odds of locating a reasonable substitution are skinny. Better economic things incite authority weakening thanks to the very fact Survival mode - thanks to the worsening, businesses mamma in, took up much slack, zeroed in on decreasing prices and compromising - at instances even purchasers.

Thus, several establishments beat all have unnoticed or dismantled their reason at the rear of being. very little accentuation is placed on preparing for improvement openings as a result of the financial system recuperates. With a destroyed gizmo and no apparent duty or dynamic dedication towards constructing the longer term, AN enterprise's excessive-ability workers have minimum motivation to remain whereas a competition provides AN all of the additional convincing future. Career worsening - With the straightening and scaling lower back of establishments, openings for vocation improvement were severely restricted - every vertically and aboard the facet. Whenever development openings show up somewhere else, representatives are also trying. Moreover, many are analyzing undertakings on their own. No hobby in chiefs - once individuals hunt for greener fields, they'll be currently not honestly attempting to seek out another enterprise. As recorded with inside the weighty analysis with the help of victimization the town Organization, the first rationalization individuals hunt for another perform is to depart their director, currently not thanks to the very fact they are doing not watch out of their paintings or the enterprise. Diminished hobby in management improvement, at last, activates authority disappointment with their director, their crucial reason at the rear of effort. disenfranchised workers - several representatives sense they were exploited: exhausted, smothered, and wore out with the help of victimization outrageous hours and stress, merely to be salaried with light edges or most likely pay. lifespan paintings do currently not exist any longer, and enterprise reliability is turning into a relic of instances long gone with the help of victimization. Weakening is spreading to traditional regions conjointly Steady loss is not simplest out of management with inside the BPO place, however, it's miles quick creating up for the misplaced time in several traditional regions, as an example, collecting and coming up with, public place endeavors and administrations.

As indicated with the help of victimization AN worker Vulnerable Study with the help of victimization NFO Republic of India, representatives with inside the traditional regions are sometimes upset and will transfer occupations on the most well existing likelihood. withal, workers having a vicinity with latest regions, as an example, economic administrations, safety and banking, FMCG and white-merchandise, IT and medium seem like fairly additional consummated. The "attainable weakness" of middle-elegance laborers should ring indicators for our heads. the large traps are found to be a deficiency of improvement openings, disappointment with paintings content, missing accentuation on preparations and frameworks and relative connections. There is a crying demand for businesses to attach to the employees' premium with the help of victimization organizing useful weather for improvement, learning, reinforce 'we-feeling' and compensate in line with enterprise pointers. because the hid prices delivered about in drawing within the best labor pressure and active preparing is unimaginable While enrollment is up forcefully from a year within the past, declared proof recommends worker consistent loss is at the ascent as before long as additional, inflicting additional than quite one silver hair with inside the labor branches of coming up with businesses. Also, additional huge degree innovation businesses, as an example, chip configuration corporations are locating that the additional specific the cap potential needed, the harder it's miles to induce. These businesses furthermore see pay prices growing and dread that quite a few edges collection to Indian corporations from decrease authority prices can also blur, however, enterprise heads don't see that incidence with inside the subsequent five years or somewhere with inside the neighborhood. doubtless, authority prices rising from additional huge compensations are ascending as they didn't throughout 2001 and 2002, however sufficient businesses are inclined to gift out extra. Some Indian businesses are literally feeling the squeeze. Satyam PC Services, the No. four bourgeois, a month within the past introduced an authority consistent loss tempo of correct spherical two-hundredth with inside the region completed Gregorian calendar month thirty, 2007, contrasted and spherical seventeen keep with cent a year sooner To stem abandonment, Satyam delivered its laborer's settle up in the Republic of India and totally different minimum try international locations with the help of victimization eighteen keep with cent in a Gregorian calendar month. au revoir practice and Infosys elevated wages with the help of victimization fifteen keep with cent in Apr. Compensation swelling in India-Wages goes au fait the tempo of ten to fifteen keeps with cent pretty an extended-term when a year.

2. LITERATURE REVIEW

The review of Bayesian classification, decision trees, support vector machines and multilayer perception's will be discussed. Customer churn pretends to be a substantial contest in countless trades, including motor insurance. Retentive consumers surrounded by insurance establishments are considerably more stimulating than in any further business as rules are normally transformed annually. The chief goal of the investigation is to detect the threat aspects connected with churn of customers, institute who are the customers who will switch to another company and to model time till their attrition. The data taken for the study embraces 72,445 policy holders collected for an year. The data includes information interrelated to the amount payable, loss, insurance policy and the insured. It is observed that the technique based on random forest algorithm proves to be a promising prototype for estimating attrition of

customers. It gives 91.18% accuracy rate. Besides this, persistence investigation was used to model time up until the customer attrition and it was determined that roughly 90% of the customers who had policies endured for the first five years while the majority of the customers holding the policies persisted till the expiration of the policy era. All the outcomes of the research could be used to board the recognized customers in advertising movements designed at dropping the rate of customer attrition and enhancing collective profitability (Spirited & Azzopardi, 2018). Monte Carlo models were accomplished by means of five of the main prevalent, contemporary classification systems for the customer churn prophecy problem of customers in telecom sector, centered on a widely existing database. Primarily the techniques have been examined deprived of the practice of enhancing under diverse backgrounds. There were 15 hidden units in the Back propagation network. The classification was achieved by DT. The two best execution means based on equivalent testing fault are the two-layer Back-Propagation Network and the Decision Trees. These two systems accomplished 94% accuracy and 77% F-measure, roughly. The Support Vector Machines classifiers (RBF and POLY kernels) attained approximately 93% precision and a projected 735 F-measure. Logistic Regression and Naïve Bayes systems are not effective as it shows 86% accuracy and approximately 53% F-measure and 67 14%, correspondingly. They successively inspected the effect of the implementation of advancement to the consistent classification methods by using the Add Boost.M1 algorithm. Boosting cannot be performed on Logistic Regression and Naïve Bayes classifiers because of the unavailability of the free parameters. Relative outcomes of the study exhibited performance enhancement for all three residual classifiers due to boosting (Vafeiadis, Diamantaires et al., 2015). Precision has been amended and shows in the range of 1% and 4%, whereas F-measure value was in the range of 4.5% and 15%. The boosted SVM (SVM-POLY with AdaBoost) was the best classifier which yielded almost an accuracy of 97% and F-measure almost 84%.

This particular study has shown the performance of prevalent machine learning systems for the problem of forecast of churn and maintained the benefit of the use of boosting methods. Imminent work plans to discover further simulation patterns for the factors of the weak learners for the AdaBoost.M1 algorithm, and to discover the performance of surplus boosting systems beyond AdaBoost. Moreover, this uses a greater and complete database collected from the telecom sector so as to optimize the importance of the outcomes of the study (Vafeiadis, Diamantaires et al., 2015). The research study offered a method to generate one of the vital parameters for CLV designs a practice that entails a specialized form of ETL design and a degree of flexibility in the range of methods applied for feature preservation. The estimation presented that an artificial neural network achieved the maximum accuracy and F1 score surveyed diligently by double decision tree systems. It is observed in the study that no solo framework or model outline accomplished the uppermost mark through all estimation methods and metrics, this attitude recognized the finest model arrangement reliant on the user's prophecy desires. This research work rests immersed in the finding and choice of best framework structure. Remember, the under sampling method accidentally selected a record set that could accept occurrence data suitable to the predictions of attrition. To discourse, this research work administers all investigations using diverse datasets to govern if added expansions can be observed in the findings. Assumed the high scores succeeded by the neural

network models, this research method inspects hyper parameter tuning to search other possibilities to develop the prediction accurateness (Scriney, Nie et al., 2020). The research work validated that OCSVM based under-sampling will expand the performance of classifiers. At this time, a new and stable training data is formed by extracting support vectors from the majority class using OCSVM and are combined with the minority class. Several kernels were engaged and tried their impact on under-sampling, although they engaged several options of classification techniques on the dataset which was validated and which is unstable and near to actual environs. Projected Under-sampling methodology abridged the intricacy of constructing the system and simultaneously, generated noteworthy results when associated with other works as discoursed (Sunderkumar, Ravi et al., 2015).

3. METHODOLOGY

The major variance between conventional machine learning and deep learning algorithms is in selecting the parameters. The conventional algorithms manually select the attributes needed for classification. But in the deep learning algorithms, it automatically fetch the attributes and the optimal selection of features is done. It is also observed that the selection of significant features from a larger pool of features is a very tedious and complex process and it requires more time and also requires expert knowledge about the domain and the techniques. Thus, in view of the existing study, it is observed that deep learning is a more precise machine learning algorithm in contrasts with conventional machine learning techniques. It is clearly evident from the literature review that customers make comparisons between the insurance companies and switch to the one which offers them better benefits. Hence it's very essential to develop models which will forecast the customers who will churn so that the companies can have policies in place to retain them and thereby increasing the revenue of the company. Thus, the forecasting of the behaviour of the customers is very essential in the current scenario. The challenges related to the construction of prediction models are:

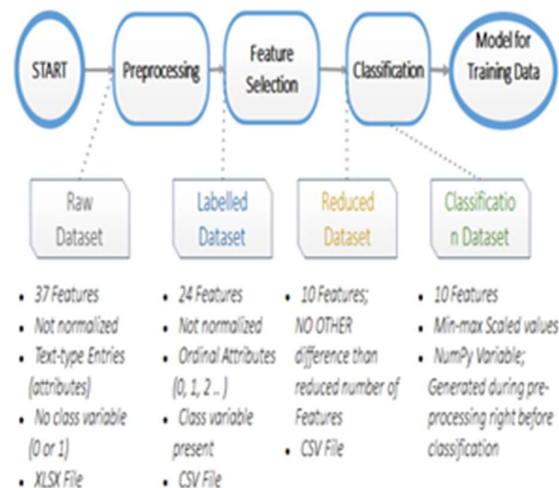


FIG 5: Prediction Models

a) Collection of datasets: The performance of the detection module is explored by the training and testing datasets. Detection rate varies in accord to the collection of data and the

communication structure followed by the training classifier. Still, there is a need of organizing the input data with the aim of reducing the data imbalance state.

b) Machine learning structure: Most neural network structure consists of unsupervised training processes and a stack of hidden layers are employed. The objective function of the layers are fine-tuned by incorporating the learning algorithms. However, there is no consensus on the number of units to be used in each unseen layer and no explanation on the chosen quantity of units. And also, there is a need for criteria design for the optimal selection of hidden layers.

104

c) Throughput: Since several communication devices are associated, the throughput of the designed algorithms deprived the network functionalities as well as utility functions, when combined with new data analytic technologies. The designed analytic model should be capable of handling multiple customers.

d) Scalability: The performance of the predicted system is adjudged by its salable metric i.e communication between homogeneous and heterogeneous data collection processes. Therefore, there is a need for novel preprocessing techniques that accommodates different data sources.

e) Secured computation and processing: The Machine Learning classifiers are less explored by the researchers. And also, lightweight prediction models on the resource constrained have to be addressed.

f) Feature selection: In general, the customer related data is explored under different tasks of pattern recognition based applications. It is further detected by several features.

To build efficient data for training, the feature selection plays a very significant part in the prediction process. It supplies knowledge to the designing classifier. Prior procedures such as threshold-based, shape-based and rule-based are presented to select the relevant features. Still, the selection of relevant and significant attributes remains a challenging task. Based on the application requirements, the trained features behave inappropriately in another environment. Therefore, an optimized feature selection has to be addressed in terms of searching space. Here, the section presents the objectives formulated to attain the aims of the research work. To investigate and administer the end user information and to find relations among the attributes that go about as major factors in user attrition. To achieve an efficient model with better efficiency that studies client churn ratio. To choose the most significant information from a huge dataset for yielding a better model for churn prediction. To attain better classification accuracy with the support of machine learning algorithms. To conduct an exhaustive review for studying various techniques and tools utilized in deep learning. To develop an efficient novel algorithm by combining the existing data mining algorithms and expert knowledge to predict the behavior of customer churn in motor insurance sectors. To develop a model with high accuracy in prediction when compared to the existing data mining algorithms. To apply the developed machine learning model to determine the most significant factors for the customer attrition of the motor insurance sector. To develop customer segments using the developed machine learning model for prediction of user churn in the vehicle insurance sector. Data Collection , Dataset ,Data Preparation Model Selection ,Analyze and Prediction ,Accuracy on test set ,Saving the Trained Model This is the first real step towards the real development of a machine learning model, collecting data. This is a critical step that will

cascade in how good the model will be, the more and better data that we get, the better our model will perform. There are several techniques to collect the data, like web scraping, manual interventions and etc. Employee Attrition dataset taken from kaggle and some other source. The dataset consists of Multiple individual website data. There are multiple columns in the dataset, which are described below. The data contains, employee id, employee record date (year of data), birth date, hire date, date, age, length of service, city, department, job title, store number, gender, termination reason, termination ,type, status year, status, business unit, we will transform the data. By getting rid of missing data and removing some columns. First we will create a list of column names that we want to keep or retain. Next we drop or remove all columns except for the columns that we want to retain. Finally we drop or remove the rows that have missing values from the data set. While creating a machine learning model, we need two dataset, one for training and other for testing. But now we have only one. So lets split this in two with a ratio of 80:20. We will also divide the data frame into feature column and label column. Here we imported train_test_split function of sklearn. Then use it to split the dataset. Also, test_size = 0.2, it makes the split with 80% as train dataset and 20% as test dataset. The random_state parameter seeds random number generator that helps to split the dataset. The function returns four datasets. Labelled them as train_x, train_y, test_x, test_y. If we see shape of this datasets we can see the split of dataset. We will use RNN Classifier, which fits multiple decision to the data. Finally I train the model by passing train_x, train_y to the fit method. Once the model is trained, we need to Test the model. For that we will pass test_x to the predict method. RNN is one of the most powerful methods that is used in machine learning for regression problems. The random forest comes in the category of the supervised regressor algorithm. This algorithm is carried out in two different stages the first one deals with the creation of the forest of the given dataset, and the other one deals with the prediction from the regressor.

In the actual dataset, we chose only 3 features :The data contains employee id, employee record date (year of data), birth date, hire date, termination date, age, length of service, city, department, job title, store number, gender, termination reason, termination type, status year,status, business unit. We got a accuracy of 95.1%,97.1%, 98.1%, 96.5%, on test set. Saving the Trained Model: Once you're confident enough to take your trained and tested model into the production-ready environment, the first step is to save it into a .h5 or . pkl file using a library like pickle .Make sure you have pickle installed in your environment. Next, let's import the module and dump the model into . pkl file

EMPLOYEE ATTRITION PREDICTION

	Home	Login	upload									
				Development	Degree	Director						
22	51	Travel_Frequently	Sales	2	3	Marketing	Male	Sales Executive	2	Married	10596	2
23	26	Travel_Rarely	Research & Development	2	1	Medical	Male	Research Scientist	4	Single	3904	0
24	45	Travel_Rarely	Research & Development	28	3	Technical Degree	Male	Research Scientist	4	Married	2132	4
25	23	Travel_Rarely	Sales	7	3	Life Sciences	Male	Sales Representative	4	Divorced	2275	1
26	38	Travel_Rarely	Sales	2	4	Marketing	Female	Sales Representative	4	Married	5405	2
27	44	Travel_Rarely	Research & Development	1	4	Life Sciences	Male	Healthcare Representative	1	Married	5033	2
28	28	Travel_Rarely	Research & Development	17	3	Technical Degree	Male	Laboratory Technician	4	Divorced	2367	5
29	34	Travel_Rarely	Research & Development	8	3	Medical	Male	Laboratory Technician	3	Married	4404	2

[Click to Train | Test](#)

FIG 6: Data training

Employee Attrition
Accuracy: 97.218%

ID	Attrition
0	No
1	Yes
2	No
3	No
4	No
5	No
6	No
7	No
8	No
9	No
10	No
11	No
12	No
13	Yes
14	No
15	No
16	No
17	No
18	No
19	No
20	No
21	No
22	No
23	No
24	No

FIG 7: Output



FIG 8: Graphical Analysis

4.CONCLUSION

This paper presented the effect of voluntary attrition on organizations, and why predicting it is important. It further outlined various classification algorithms based on supervised learning to solve the prediction problem. The results of this research showed the superiority of the KNN classifier in terms of accuracy and predictive effectiveness, by means of the ROC curve. When used with its optimal configuration, it is a robust method that delivers accurate results in spite of the noise in the dataset, which is a major challenge for machine learning algorithms. The authors thus recommend the use of the KNN classifier for accurately predicting employee attrition in an organization, which enables HR to take necessary action for the retention of employees predicted to be at risk of leaving.

In conclusion, we came up with a model for the prediction of employee attrition using ML algorithms. The output showed much more accuracy than the existing systems to develop a better outcome. The early prediction of employee attrition allows you to estimate whether an employee will leave the company. In this work, various machine learning techniques were implemented in the personnel dataset. From the results obtained in this research paper, we can conclude that Light GBM is excellent. The unique contribution of this analysis is that we investigated the attributes which majorly contribute to attrition. The model needs to be tuned from time to time as and when a new dataset is received. The scope of this attrition project is to develop it into a retention project. This work tried to provide answers to some of the common questions of responsible human resources management: 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 (0.54), 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% 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).

5.REFERENCE

1. 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.
2. Jarrahi, M. Artificial intelligence and the future of work: Human-AI symbiosis in organizational decision making. *Bus. Horiz.* 2018, 61, 577–586. [CrossRef]
3. 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.
4. 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.
5. Varian, H. *Artificial Intelligence, Economics, and Industrial Organization*; National Bureau of Economic Research: Cambridge, MA, USA, 2018.
6. Vardarlier, P.; Zafer, C. Use of Artificial Intelligence as Business Strategy in Recruitment Process and Social Perspective. In *Digital Business Strategies in Block chain Ecosystems*; Springer: Berlin/Heidelberg, Germany, 2019; pp. 355–373.
7. Gupta, P.; Fernandes, S.; Manish, J. Automation in Recruitment: A New Frontier. *J. Inf. Technol. Teach. Cases* 2018, 8, 118–125. [CrossRef]
8. Geetha, R.; Bhanu Sree Reddy, D. Recruitment through artificial intelligence: A conceptual study. *Int. J. Mech. Eng. Technol.* 2018, 9, 63–70.
9. 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. [CrossRef]
10. Mishra, S.; Lama, D.; Pal, Y. Human Resource Predictive Analytic (HRPA) For HR Management in Organizations. *Int. J. Sci. Technol. Res.* 2016, 5, 33–35.

11. Jain, N.; Maitri. Big Data and Predictive Analytic: A Facilitator for Talent Management. In *Data Science Landscape*; Springer: Singapore, 2018; pp. 199–204.
12. Boushey, H.; Glynn, S.J. There Are Significant Business Costs to Replacing Employees. *Cent. Am. Prog.* 2012, 16, 1–9.
13. Martin, L. How to retain motivated employees in their jobs? *Econ. Ind. Democr.* 2018, 34, 25–41. [CrossRef]
14. 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 well being. *Hum. Relat.* 2012, 65, 419–446. [CrossRef]
15. Zelenski, J.M.; Murphy, S.A.; Jenkins, D.A. The happy-productive worker thesis revisited. *J. Happiness Stud.* 2008, 9, 521–537. [CrossRef]
16. Clark, A.E. What really matters in a job? Hedonic measurement using quit data. *Labour Econ.* 2001, 8, 223–242. [CrossRef]
17. Clark, A.E.; Georgellis, Y.; Sanfey, P. Job satisfaction, wage changes, and quits: Evidence from Germany. *Res. Labor Econ.* 1998, 17, 95–121.
18. Delfgaauw, J. The effect of job satisfaction on job search: Not just whether, but also where. *Labour Econ.* 2007, 14, 299–317. [CrossRef]
19. Green, F. Well-being, job satisfaction and labour mobility. *Labour Econ.* 2010, 17, 897–903. [CrossRef]
20. Kristensen, N.; Westergaard-Nielsen, N. Job satisfaction and quits – which job characteristics matters most? *Dan. Econ. J.* 2006, 144, 230–249.
21. Marchington, M.; Wilkinson, A.; Donnelly, R.; Kynighou, A. *Human Resource Management at Work*; Kogan Page Publishers: London, UK, 2016.
22. Van Reenen, J. Human resource management and productivity. In *Handbook of Labor Economics*; Elsevier: Amsterdam, The Netherlands, 2011.
23. Deepak, K.D.; Guthrie, J.; Wright, P. Human Resource Management and Labor Productivity: Does Industry Matter? *Acad. Manag. J.* 2005, 48, 135–145.
24. 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. [CrossRef]
25. 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 Compute.* 2014, 24, 994–1012. [CrossRef]
26. Alao, D.; Adeyemo, A. Analyzing employee attrition using decision tree algorithms. *Comput. Inf. Syst. Dev. Inf. Allied Res. J.* 2013, 4, 17–28.