Retention Is All You Need

Karishma Mohiuddin
University of Bonn
Bonn, Germany
natz.karishma@gmail.com

Mirza Ariful Alam
University of Rajshahi
Rajshahi, Bangladesh
mirza.hridoy@gmail.com

Mirza Mohtashim Alam
FIZ Karlsruhe
Karlsruhe, Germany
turzo.mohtasim@gmail.com

Pascal Welke
TU Wien
Vienna, Austria
pascal.welke@tuwien.ac.at

Michael Martin
InfAI
Leipzig, Germany
martin@infai.org

Jens Lehmann*
TU Dresden, InfAI, Amazon
Dresden, Germany
jens.lehmann@tu-dresden.de

Sahar Vahdati
InfAI
Dresden, Germany
vahdati@infai.org

ABSTRACT

Skilled employees are the most important pillars of an organization. Despite this, most organizations face high attrition and turnover rates. While several machine learning models have been developed to analyze attrition and its causal factors, the interpretations of those models remain opaque. In this paper, we propose the HR-DSS approach, which stands for Human Resource (HR) Decision Support System, and uses explainable AI for employee attrition problems. The system is designed to assist HR departments in interpreting the predictions provided by machine learning models. In our experiments, we employ eight machine learning models to provide predictions. We further process the results achieved by the best-performing model by the SHAP explainability process and use the SHAP values to generate natural language explanations which can be valuable for HR. Furthermore, using "What-if-analysis", we aim to observe plausible causes for attrition of an individual employee. The results show that by adjusting the specific dominant features of each individual, employee attrition can turn into employee retention through informative business decisions.

KEYWORDS

Business Intelligence, Decision Support System, Interpretable prediction, Explainable AI, Employee Attrition and Retention, Machine Learning Models, Natural Language Generation

https://doi.org/10.1145/3583780.3615497

---

*Work done outside of Amazon.

1 INTRODUCTION

Attrition (voluntary leave of employees) and turnover (employee replacement) are among the major challenges of any business. Employee retention is crucial for reducing training and recruitment costs while preserving talent and internal knowledge for any businesses [21]. The job crisis sparked by the COVID-19 pandemic serves as an additional reminder on the significance of employee attrition analysis [20]. To retain employees, companies should emphasize on the causal factors such as salary, promotion, work environment, job satisfaction, and stock option beforehand. Current machine learning techniques for analyzing attrition and its causal determinants often rely on black-box approaches, i.e. their results are not easily interpretable by human resource experts.

We addressed this practical issue by combining simple ML models with explainable AI (XAI) techniques aiming to reduce barriers for adoption in practice. During the data preparation phase, we detected and removed outliers, fixed class imbalance and added weights on specific features. In the data analysis phase, we conducted a predictive analysis using eight Machine Learning (ML) models. Finally, we employed the XAI library SHAP (SHapley Additive exPlanations) [14, 15, 17] on the ML model that achieves the highest accuracy. Further, for the output predictions provided by the best model (in our case XGB), we gathered the correctly predicted instances and incorrectly predicted instances. Since we are focusing on the application for human resources (HR), we have integrated a natural language generation module using OpenAI to output explanations in natural language. The whole pipeline is facilitated with an explainer dashboard to further interpret the features per each employee. The specific causal factors of attrition corresponding to individual employees can be understood and adjusted by 'What-if-analysis'. The explanation provided in natural language can assist HR in a more critical decision-making process for retention policy.

2 RELATED WORK

Different ML models have been employed to predict the importance of features in attrition [3, 28]. XGBoost [5] is one of the predictive
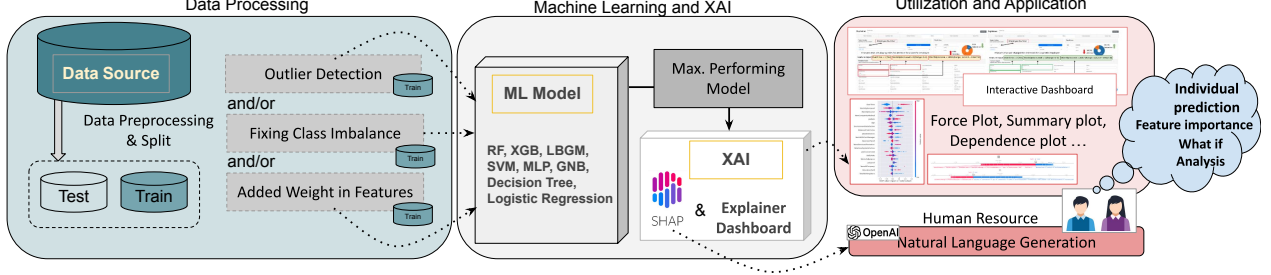
Figure 1: The workflow of the proposed decision support system (HR-DSS).

models that is used in several studies [1, 28]. In [1], XGBoost outperforms other ML models for attrition prediction. Other works use classical approaches to predict the causal factors of the attrition. In[23] the highest cross validation accuracy of 85.12% is achieved by a random forest [13] among GNB, KNN, SVM, Decision Tree, Logistic Regression. Features, such as monthly income, age, and daily rate, are suggested to be the key factors of employee turnover. Such approaches statistically evaluate and learn from historical data, without providing any explanation for the predictions. SHAP (SHapley Additive exPlanations) [14, 15, 17] is a game theoretic approach to explain the output of any machine learning model. SHAP is widely used in several domains including attrition analysis [24]. Recently, [25] applied the SHAP library [16, 25] and LIME [6] in order to interpret the predictions of ML models. However, these results did not consider the class imbalance issue typically present in employee attrition settings. We will further explore this in the paper. To build a decision support system, we have employed an outlier detection and assigned weights to important features to achieve a higher accuracy of predictions.

3 APPROACH

In this section, we present the design and development of our approach. The proposed decision support system (DSS) is called HR-DSS and uses a SHAP library to provide explainability on top of the predictive models that are used for learning. In order to explain the SHAP values in natural language, we have used the OpenAI GPT-4[1] completion library. The SHAP values for the best performing ML model are constructed as prompts with the conditions and guidelines to obtain the explanation in natural language. We deployed an XAI-based dashboard [7] that the HR personnel can observe, interact, and understand attrition prediction transparently and confidently. Figure 1 gives an overview of the proposed system. The interactive decision support system is based on the achieved predictions and serves as a reliable decision-support system.

We use the IBM HR Analytics Attrition Dataset[2], which is a widely used synthetic dataset. The dataset contains 34 features related to attrition with a sample size of 1470 records without any missing values. The dataset consists of both numerical and categorical data.

Obtaining Maximum Performing Model. The workflow starts with a list of machine learning models performing the prediction of prominent features causing attrition. As shown in Algorithm 1, let

[1]https://openai.com/gpt-4
[2]https://www.kaggle.com/datasets/pavansubhasht/ibm-hr-analytics-attrition-dataset

---

Algorithm 1: Obtain Maximum Performing Model

input : ML Models $\mathcal{MLf} = (M\_1, M\_2, ...M\_8)$
output: Max_Accuracy α_{MA} Max_perf_model M_{pm}, best_fitted_model $clf_{M_{pm}}$

1   $D \leftarrow read.data(datapath)$
2   $cat\_dict, data \leftarrow data\_process\_for\_ML(D)$
3   $x\_train, y\_train, x\_test, y\_test \leftarrow traintestsplit(data)$
4   **if** $outlier\_detect == True$ **then**
5     $(x\_train, y\_train) = remove\_outlier(x\_train, y\_train)$
6   **if** $Imblearn == True$ **then**
7     $(x\_train, y\_train) = ImbLearn(x\_train, y\_train)$
8   **if** $weighted\_feature == True$ **then**
9     $(x\_train, y\_train) = added\_weighted\_feature(x\_train, y\_train)$
10   $\alpha_{MA}, M_{pm} \leftarrow 0, None$
11   **for** $model\ in\ \mathcal{MLf}$ **do**
12     $clf \leftarrow model.initialize(args)$
13     $\beta_{MA} = crossvalscore(clf, x\_train, y\_train)$
14   **if** $\beta_{MA} > \alpha_{MA}$ **then**
15       $\alpha_{MA} \leftarrow \beta_{MA}$
16       $\mathbf{M_{pm}} \leftarrow model$
17   **end**
18   $clf_{M_{pm}} \leftarrow model.initialize(args, M_{pm})$
19   $clf_{M_{pm}}.train(x\_train, y\_train)$

$\mathcal{MLf}$ be the collection of eight ML models ($M\_1, M\_2...M\_8$) to be utilized (see Table 1). We aim at obtaining the best performing ML model M_{pm} and collecting its maximum accuracy α_{MA}. At the beginning of the Algorithm 1, the data is read and pre-processed, and unnecessary columns (i.e., Over18, EmployeeCount, EmployeeNumber, StandardHours) were removed through a combination of manual and statistical verification. Then, the data was split into the train (80%) and test (20%) sets. Depending on the parameters, outlier detection and removal, class balancing, and feature weighting are applied to the training set (see Section 4). Three distinct parameters are considered to represent these. The feature engineering techniques are broadly discussed in Section 4. We conduct 9-fold cross-validation to identify the optimal model from $\mathcal{MLf}$. Subsequently, we set up and retrain the classifier $clf_{M_{pm}}$ employing the top-performing model M_{pm} on the entire training dataset.

Single Prediction Model Explainability. To understand the prediction of the best performing ML model, SHAP values are

Algorithm 2: Generate a natural language explanation for a single prediction

input : max performing model $clf_{M_{pm}}$, feature names (`feature_names`), prediction for the instance (`prediction`), specific_employee

output: Response in Natural Language

1 shap_values ← build_explainer_on_classifier($clf_{M_{pm}}$, specific_employee);

2 shap_values_current ← $shap\_values(specific\_employee)$;

3 feature_names ← columns of X_train;

4 prompt ← build_prompt (shap_values_current, the company's rules);

5 **foreach** feature_name, shap_value *in* (feature_names, shap_values_single) **do**

6   prompt ← construct(prompt, f"The SHAP value for feature '{feature_name}' is {shap_value:.2f}.");

7 **end**

8 prompt ← construct(prompt, "Please write the reason of stay or leave in bullet points?. Also provide suggestions how to retain the employee in bullet points") ;

9 response ← OpenAI API call with prompt;

10 Return response;

Algorithm 3: XAI model for single prediction

input : $clf_{M_{pm}}$, specific_employee

output: Force plot

1 $AnalyzeSingle\_pred(specific\_employee, clf_{M_{pm}})$

2 $V_a ← shap_{fp}(specific\_employee)$

3 return V_a

Algorithm 4: XAI model for overall prediction

input : $M_{pm}, specific\_employee, x\_test$

output: Summary plot & Explainer Dashboard

1 $AnalyzeFull\_pred(clf_{M_{pm}}, x\_test)$;

2 $V_a ← shap_{sp}(M_{pm}.shap\_value[1], features\_names)$

3 $V_b ← shap_{fp}(x\_test, clf_{M_{pm}}.explainer.expected\_value[1], shap\_value[1])$

4 $explainer_{db} ← Initialize(clf_{M_{pm}}, x\_test, cat\_dict)$

5 $explainer_{db}.run()$

---

initialized. To interpret the model output clearly, we obtain the overall prediction (of the label Attrition) of the model on a particular instance of the test data. In Algorithm 3, the maximum performing model α_{MA}, as provided by Algorithm 1 is taken as the input and is used to predict the label of a single example from the test set or it can be a specific employee instance.

In this part, to analyze the single prediction, several parameters are taken as arguments. A particular employee from the test set or a new specific instance having the same features. We put the test features' specific_employee in the visualization object V_a of the SHAP library. Further, we also applied SHAP values to provide explanations in natural language. In this regard, we are using the OpenAI completion library wrapped with the model "text-davinci-003" [18]. The implementation of natural language generation is shown in Algorithm 2.

To illustrate the dependence plot, several arguments are considered as the function parameter. In this case, we take the maximum performing model's SHAP value with index 1, test features of a specific employee, and pass it to the variable V_b for the purpose of visualization. To initialize the explainer dashboard with the test data. The goal of using the explainer dashboard is to explain the prediction including what if analysis.

In the end, the dashboard is executed to analyze and have a proper observation of the interactive explanation. For single prediction, the XAI algorithm is described in Algorithm 3.

Global Prediction Model Explainabilty. In Algorithm 4, which is deployed to analyze the prediction of the full population, α_{MA}, specific_employee and x_test are taken as the input parameters for the function. Indices of x_test are obtained as mentioned above. To visualize the features of one empployee, specific_employee is

needed. V_a is taken into consideration to be able to visualize the summary plot with SHAP. In this regard, the maximum performing model's Shap values and the feature names are provided. Moreover, to demonstrate the cumulative force plot, we take the expected values of maximum performing model from the explainer and the corresponding SHAP values for visualization.

An explainer dashboard is used to provide further analysis. The predicted indices of the test data, the maximum performing model $clf_{M_{pm}}$ and the categorical dictionary are provided in order to initialize the explainer object $explainer_{db}$. Analysing the wrong predictions by the model is also possible. However, it is noteworthy that the use of this dashboard is not to provide predictions, but to explain possible causes for HR, after an employee leaves.

4 DATA PROCESSING

We now describe the components of our data processing pipeline in more detail. The IBM HR Analytics Attrition Dataset[3], contains 34 features related to attrition with a sample size of 1470 records without any missing values. To achieve good predictive performance with simple machine learning models we turn to classic data pre-processing techniques. Due to the privacy policy and unavailability of real public dataset of company, we took this synthetic data.

4.1 Class Imbalance for Learning Models

The dataset has a class imbalance problem meaning that the total number of positive examples is much smaller than the number of negative examples. Therefore, re-sampling the data is essential to increase the importance of positive examples. The dataset contains 237 examples labeled "Yes" for Attrition and 1233 examples labeled "No" for Attrition. Undersampling of the negative examples leads to loss of instances from the data that may hold important information. To avoid this, several oversampling techniques have been employed, namely SMOTE [4], ADASYN [11], as well as SMOTE + Tomek [22]), which is a hybrid version of SMOTE.

---

[3]https://www.kaggle.com/datasets/pavansubhasht/ibm-hr-analytics-attrition-dataset

Kahinsuddin, Mirza Ariful Alam, Mirza Mohtashim Alam, Pascal Welke, Michael Martin, Jens Lehmann, and Sahar Vahdati

ML Algorithm	Acc. F1 Score	Imbalanced	SMOTE	ADASYN	SMOTE+Tomek
Random Forest	Acc.	87.76%	87.07%	87.76%	87.76%
	F1 S.	18.18%	32.14%	28.00%	35.09%
Decision Tree	Acc.	78.23%	79.59%	77.89%	78.23%
	F1 S.	17.95%	33.33%	30.70%	25.58%
GNB	Acc.	84.01%	57.48%	59.18%	58.16%
	F1 S.	47.19%	28.77%	29.41%	29.71%
Logistic Regression	Acc.	84.69%	64.63%	60.88%	64.97%
	F1 S.	4.25%	28.77%	25.81%	27.97%
MLP	Acc.	85.03%	61.90%	31.97%	42.86%
	F1 S.	0.00%	25.33%	25.37%	26.32%
LGBM Classifier	Acc.	88.44%	88.44%	87.76%	87.41%
	F1 S.	37.04%	41.38%	35.71%	37.29%
SVM	Acc.	86.73%	86.73%	13.27%	86.73%
	F1 S.	0.0%	0.0%	23.42%	20.00%
XGB	Acc.	88.44%	85.37%	87.76%	86.73%
	F1 S.	35.29%	29.51%	37.93%	31.58%

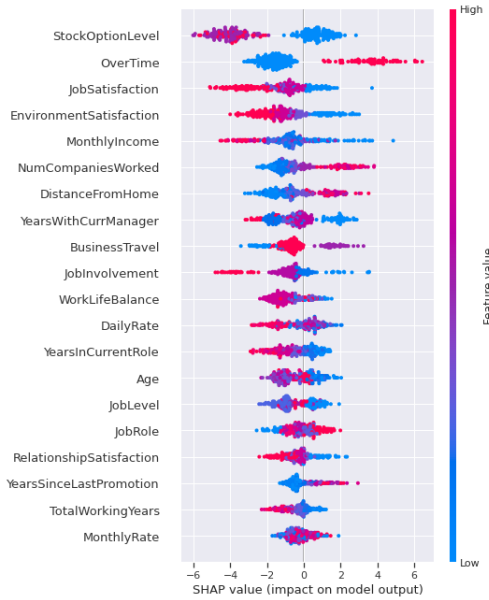**Table 1: Accuracy comparison among ML models trained on imbalanced data and balanced data.**

Figure 2: LGBM after balancing the data with SMOTE.

These techniques are employed on the selected ML models, namely random forest, XGB, MLP, LGBM, decision tree, logistic regression, SVM, and Gaussian Naive Bayes. A fixed seed is used to avoid the problem of result deviation in the performance. However, the accuracy of the balanced variant was then reduced for most of the ML models. The comparison among imbalanced and balanced variant concerning F1 scores is demonstrated in Table 1. LGBM showed best performance of 88.44% with SMOTE among all the ML models, LGBM with SMOTE also returned a high F1 score. In Figure 2, it can be observed that there are more overlapping data points for medium (purple color) to high (red color) feature values of a particular feature. For example, medium to high values of Stock-OptionLevel are tightly mixed without being separable and many of the positive classes are shifted on the negative side in this case. The overlapping points of data in visualization may cause difficulties in
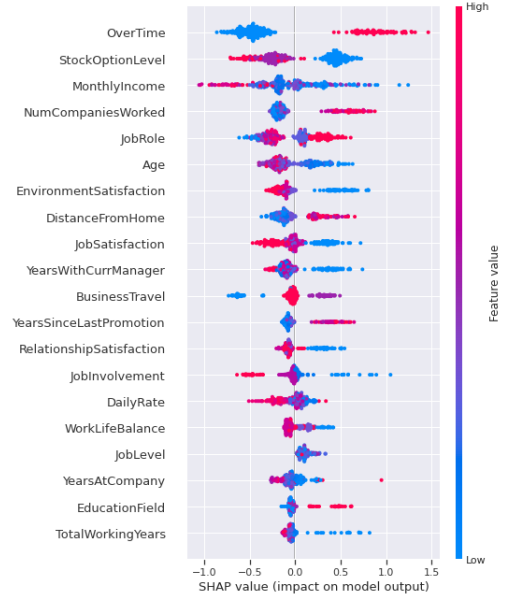

Figure 3: XGB summary plot using added weight only

understanding which types of employees would actually determine to leave the company. The best model is expected to provide the instances of positive and negative classes separated by the vertical line in the middle of the visualizations of the summary plots that divide the class labels. Moreover, by balancing the data there are not many improvements compared to the original data. Aligned with our observation, other works [2, 8, 26] also mention that accuracy measures are not reliable for imbalance classes, which can be due to severe class imbalance often mimicking a high accuracy.

4.2 Outlier Detection and Weighted Features

To improve the overall predictions of the ML models, an outlier detection technique has been introduced using Isolation Forest. [12] Isolation Forest has the same essence as the Decision tree algorithm. After providing a set of features, the algorithm randomly selects a feature from those features to remove outliers and then select a random split value to separate the maximum and minimum value of that feature. Besides, weights are added to StockOptionLevel and JobLevel features based on business theories [19] [10] to create a more separable group of data points because according to the theory, these two features consider as important factors of attrition. In this approach, after splitting the train and test data, Isolation forest outlier detection and weighted features are considered together to improve the performance. Additionally, the outlier detection technique and the weighted features are also experimented separately to compare the performance. According to Table 2, XGB has achieve improvements with improved F1 scores for every model relative to the original model. XGB with only added weights on the mentioned specific features has achieved the best result with 89.12% accuracy. We have chosen the XGB model for interpretation with XAI for this phase to evaluate with summary plot.

ML Algorithm	Accuracy and F1 Score	Original Data	With Outlier	With Weights	With Outlier and Weights
Random Forest	Acc.	87.76%	87.76%	87.07%	88.44%
	F1 S.	18.18%	14.29%	17.39%	22.72%
Decision Tree	Acc.	78.23%	77.89%	78.57%	40.14%
	F1 S.	17.95%	34.88%	22.22%	20.72%
GNB	Acc.	84.01%	82.31%	84.01%	81.97%
	F1 S.	47.19%	47.19%	47.19%	46.46%
Logistic Regression	Acc.	84.69%	86.39%	86.36%	86.73%
	F1 S.	4.26%	0.00%	4.76%	4.88%
MLP	Acc.	85.03%	86.05%	86.39%	86.73%
	F1 S.	0.00%	0.00%	0.00%	0.00%
LGBM Classifier	Acc.	88.44%	88.10%	87.07%	86.39%
	F1 S.	37.04%	38.60%	38.71%	35.48%
SVM	Acc.	86.73%	86.73%	13.27%	86.73%
	F1 S.	0.00%	0.00%	0.00%	0.00%
XGB	Acc.	88.44%	88.78%	89.12%	88.78%
	F1 S.	35.29%	44.83%	40.00%	40.00%

Table 2: Accuracy comparison among ML algorithms trained on original data, data with added weighted features and outlier removal techniques, and a combination of both methods

The accuracy of XGB improves with better F1 scores for features with added weights. Additionally, for some of the features like YearsWithCurrentManager, JobRole data points are more distinguishable in Figure 3. For XGB with weighted feature, there are less wrong predictions (32), while 262 are correctly predicted instances. Hence, XGB has been chosen for further interpretation as it is the best performance model achieving 89.12% accuracy and its F1 Score as mentioned in Table 2.

To make the explainable more understandable, NLG applied and the explainer dashboard[4] has been adapted on top of SHAP to create an interactive web-based GUI platform to interpret and analyze the predictions of the model further for retention policy.

5 RESULT AND ANALYSIS

The XGB model has demonstrated better accuracy of 89.12% than other ML models for this data and has faster training speed with low resources [5]. Hence, the following evaluations by SHAP visualizations are on XGB.

5.1 SHAP Summary Plot

The summary plot is the combination of feature importance and effect. In summary plots, the x-axis indicates the Shapley values, and the y-axis represents the features. The feature values are arranged with color intensity between low (blue) to high (red). The summary plot in Figure 4 is a global-level visualization: Consider the 'OverTime' feature. The blue points indicate the employees do not work overtime (red dots shows the opposite). Hence, employees with no overtime would remain in the organization and shift towards the negative Shap value (left side), whereas the red dots mean employees with working overtime are more likely to make a positive decision towards attrition.

5.2 SHAP Force Plot

To analyze an individual prediction, the force plot can be a useful module for understanding unique reasoning. For example, the
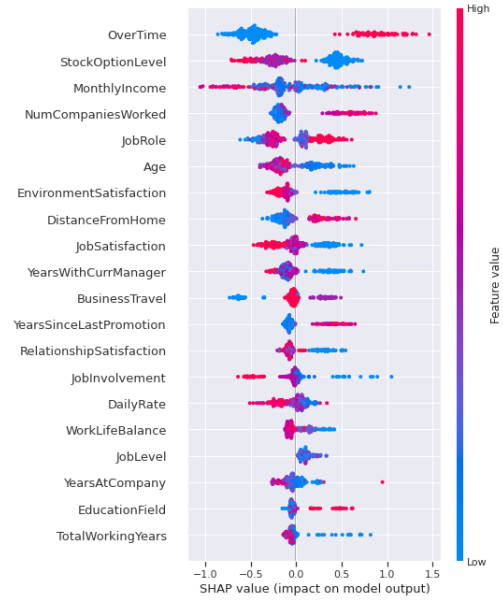
---

[4]https://explainerdashboard.readthedocs.io/en/latest/

Figure 4: Summary Plot for overall prediction.

upper side of Figure 5 indicates that the person aged 19, a female, is strongly adamant about her attrition decision. The red color indicates the dominant factors for the attrition where the widest bars like over time, low monthly income, age, low environmental satisfaction, no stock option plan, etc., have greater influences in her decision. This visualization confirms that younger employees are prone to attrition than senior employees because of having moderate monthly income, working overtime and receive low stock options. [9][27].On the contrary, in the bottom force plot, the SHAP output predicts and demonstrates that a male employee, aged 37, would most likely stay in the organization based on the factors of no overtime, job involvement, job satisfaction, higher monthly income, etc.

5.3 Natural Language Generation Based on SHAP Values

We have used the Shap value output to generate natural language as described in Algorithm 2. The prompt for the OpenAI completion module is constructed in such a way that it contains the description what the Shap values contribute to the outcome. We have also added some dummy general policy of the company as a part of the prompt i.e, company does not offer remote work, less stock option for young employees etc. An example LLM output looks as follows:

Based on the SHAP values for this (woman, 19 years old) instance, the machine learning model has predicted that the employee will leave the company. Factors that contributed to this decision include:

- The employee is young age (SHAP value: 0.46)
- Their monthly income (SHAP value: 1.90)
- Their level of stock option (SHAP value: 0.53)
- Their years with current manager (SHAP value: 0.60)
- Their overtime hours (SHAP value: 3.01)
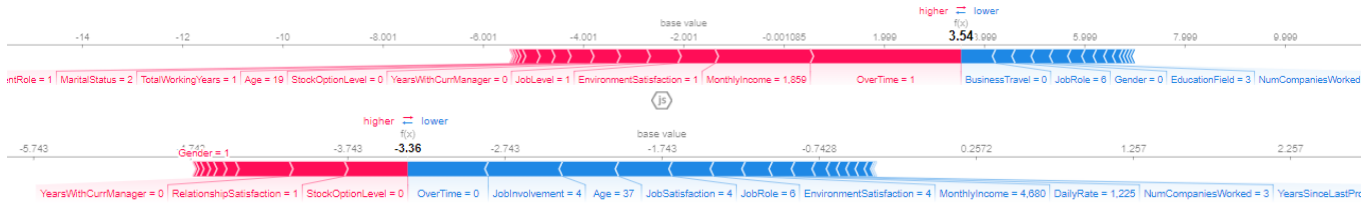- The employees job level (SHAP value: 0.62)

Figure 5: Force plot for individual female attrition decision (top) vs. male decision to stay (bottom)
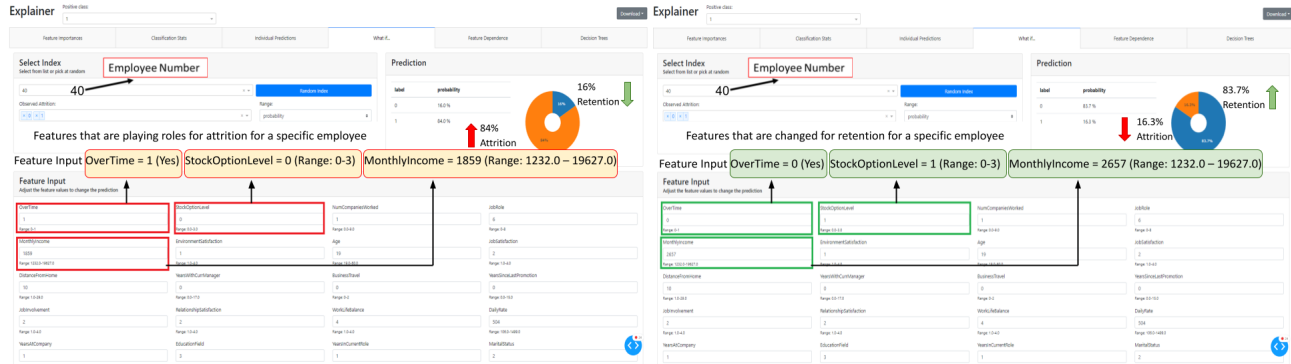

Figure 6: Comparison between before and after using What-if-analysis for an individual employee who has probability of attrition 84%. After using the What-if-analysis to change particular features for the employee, the attrition rate decreased to 16.3% and the retention ratio increased to 83.7%

In order to retain the employee, the company can:

- Increase the employee's salary and level of stock option
- Increase job opportunities and promotions that are based on performance rather than years
- Offer part-time options and flexible work hours
- Offer more remote work options
- Provide additional paid leaves
- Provide more remuneration when working extra hours

5.4 Interactive XAI Dashboard

To retain talented and resourceful employees, the business community demands transparency and reliability of ML models. To ensure this demand, the same SHAP values are integrated to an interactive explainer dashboard. In Figure 6, an individual employee is visualised with a predicted 86% chance of leaving the organization with some key attributes. From both the force plot and the interactive dashboard, it has been observed that low monthly income, overtime, and no stock option level are the most significant criteria for plausible attrition. This has been explained directly with natural language generation. The What-if-analysis module assists HR by analyzing the probable retention of a specific employee by increasing and decreasing the causal factors and let HR to be aware of retention policy. With the increment of monthly salary and increased stock option level while dissolving the over time issue, the chance of attrition of this specific employee will decrease to 16.3%. Due to the page constraint, each of the parameters effects are not shown.

6 CONCLUSION

In this work, we investigated eight machine-learning approaches for the application of HR data for attrition and retention. Among them, the best performing model is XGB with 89.12% accuracy in providing predictions on the considered benchmark data. Further, specific features like OverTime, and StockOptionLevel, MonthlyIncome have been determined by model as top contributing reasons to attrition. The proposed approach, named HR-DSS, is designed as an XAI-driven decision support system with an interactive user interface. To further enhance the decision support system, we generated natural language from the SHAP values for the explanation to human resources. This work shows the impact of explainable machine learning algorithms on employee attrition,retention policy and decision-making. An immediate next step for this work is to include the NLG in the dashboard and enable the investigation of other neural network-based as well as deep-learning models. To overcome ethical issue and hallucination of Open-AI, we will only use the policy/guidelines documentation of an organization by applying prompt engineering. Hence, the LLM will generate results according to the XAI SHAP values and address the issue with advice from the guidelines of that organization. The HR-DSS will be an end-to-end AI-based intelligent system aiming at assisting human resource units in different organizations to retain highly skilled employees and measure the performance of all employees with trustworthy justifications based on explainable AI.

REFERENCES

[1] Pankaj Ajit. 2016. Prediction of employee turnover in organizations using machine learning algorithms. *algorithms* 4, 5 (2016), C5.

[2] Qasem A Al-Radaideh and Eman Al Nagi. 2012. Using data mining techniques to build a classification model for predicting employees performance. *International Journal of Advanced Computer Science and Applications* 3, 2 (2012).

[3] Sarah S Alduayj and Kashif Rajpoot. 2018. Predicting employee attrition using machine learning. In *2018 international conference on innovations in information technology (iit)*. IEEE, 93–98.

[4] Nitesh V Chawla, Kevin W Bowyer, Lawrence O Hall, and W Philip Kegelmeyer. 2002. SMOTE: synthetic minority over-sampling technique. *Journal of artificial intelligence research* 16 (2002), 321–357.

[5] Tianqi Chen and Carlos Guestrin. 2016. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*. 785–794.

[6] Jürgen Dieber and Sabrina Kirrane. 2020. Why model why? Assessing the strengths and limitations of LIME. *arXiv preprint arXiv:2012.00093* (2020).

[7] Oege Dijk. 2020. explainerdashboard. https://github.com/oegedijk/explainerdashboard.

[8] Amir Mohammad Esmaieeli Sikaroudi, Rouzbeh Ghousi, and Ali Sikaroudi. 2015. A data mining approach to employee turnover prediction (case study: Arak automotive parts manufacturing). *Journal of industrial and systems engineering* 8, 4 (2015), 106–121.

[9] Francesca Fallucchi, Marco Coladangelo, Romeo Giuliano, and Ernesto William De Luca. 2020. Predicting employee attrition using machine learning techniques. *Computers* 9, 4 (2020), 86.

[10] Yanjun Guan, Yueran Wen, Sylvia Xiaohua Chen, Haiyang Liu, Wei Si, Yuhan Liu, Yanan Wang, Ruchunyi Fu, Yuyan Zhang, and Zhilin Dong. 2014. When do salary and job level predict career satisfaction and turnover intention among Chinese managers? The role of perceived organizational career management and career anchor. *European Journal of Work and Organizational Psychology* 23, 4 (2014), 596–607.

[11] Haibo He, Yang Bai, Edwardo A Garcia, and Shutao Li. 2008. ADASYN: Adaptive synthetic sampling approach for imbalanced learning. In *2008 IEEE international joint conference on neural networks (IEEE world congress on computational intelligence)*. IEEE, 1322–1328.

[12] Fei Tony Liu, Kai Ming Ting, and Zhi-Hua Zhou. 2008. Isolation forest. In *2008 eighth ieee international conference on data mining*. IEEE, 413–422.

[13] Yanli Liu, Yourong Wang, and Jian Zhang. 2012. New machine learning algorithm: Random forest. In *International Conference on Information Computing and Applications*. Springer, 246–252.

[14] Scott M. Lundberg, Gabriel Erion, Hugh Chen, Alex DeGrave, Jordan M. Prutkin, Bala Nair, Ronit Katz, Jonathan Himmelfarb, Nisha Bansal, and Su-In Lee. 2020. From local explanations to global understanding with explainable AI for trees. *Nature Machine Intelligence* 2, 1 (2020), 2522–5839.

[15] Scott M Lundberg and Su-In Lee. 2017. A Unified Approach to Interpreting Model Predictions. In *Advances in Neural Information Processing Systems 30*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (Eds.). Curran Associates, Inc., 4765–4774. http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf

[16] Scott M Lundberg and Su-In Lee. 2017. A unified approach to interpreting model predictions. *Advances in neural information processing systems* 30 (2017).

[17] Scott M Lundberg, Bala Nair, Monica S Vavilala, Mayumi Horibe, Michael J Eisses, Trevor Adams, David E Liston, Daniel King-Wai Low, Shu-Fang Newman, Jerry Kim, et al. 2018. Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. *Nature Biomedical Engineering* 2, 10 (2018), 749.

[18] John J Nay. 2023. Large Language Models as Corporate Lobbyists. *arXiv preprint arXiv:2301.01181* (2023).

[19] Paul Oyer and Scott Schaefer. 2005. Why do some firms give stock options to all employees?: An empirical examination of alternative theories. *Journal of financial Economics* 76, 1 (2005), 99–133.

[20] Kim Parker, Juliana Menasce Horowitz, and Rachel Minkin. 2022. Covid-19 pandemic continues to reshape work in America. https://www.pewresearch.org/social-trends/2022/02/16/covid-19-pandemic-continues-to-reshape-work-in-america/

[21] Jack J Phillips and Adele O Connell. 2003. *Managing employee retention: a strategic accountability approach*. Routledge.

[22] Ronaldo C Prati, Gustavo EAPA Batista, and Maria Carolina Monard. 2004. Learning with class skews and small disjuncts. In *Brazilian Symposium on Artificial Intelligence*. Springer, 296–306.

[23] Madara Pratt, Mohcine Boudhane, and Sarma Cakula. 2021. Employee Attrition Estimation Using Random Forest Algorithm. *Baltic Journal of Modern Computing* 9, 1 (2021), 49–66.

[24] Navya Sabbineni. 2020. Understanding Employee Attrition Using Explainable AI. (2020).

[25] Karthik Sekaran and S Shanmugam. 2022. Interpreting the Factors of Employee Attrition using Explainable AI. In *2022 International Conference on Decision Aid Sciences and Applications (DASA)*. IEEE, 932–936.

[26] Randall S Sexton, Shannon McMurtrey, Joanna O Michalopoulos, and Angela M Smith. 2005. Employee turnover: a neural network solution. *Computers & Operations Research* 32, 10 (2005), 2635–2651.

[27] Devesh Kumar Srivastava and Pradeep Kumar Tiwari. 2020. An analysis report to reduce the employee attrition within organizations. *Journal of Discrete Mathematical Sciences and Cryptography* 23, 2 (2020), 337–348.

[28] Yue Zhao, Maciej K Hryniewicki, Francesca Cheng, Boyang Fu, and Xiaoyu Zhu. 2018. Employee turnover prediction with machine learning: A reliable approach. In *Proceedings of SAI intelligent systems conference*. Springer, 737–758.