. **Paper No: SC-16** **Smart Computing**

What makes job satisfaction in the information technology industry?

Nimasha Arambepola*
Software Engineering Teaching Unit
Faculty of Science, University of Kelaniya, Sri Lanka
nimasha_2019@kln.ac.lk

Lankeshwara Munasinghe
Software Engineering Teaching Unit
Faculty of Science, University of Kelaniya, Sri Lanka
lankesh@kln.ac.lk

***Abstract* - Having a rich human resource is critical for an organization to move towards success. Especially, for business organizations such as technology companies, the human resource is the driving factor of the company's growth which depends on employees' motivation, skills and quality of work. Employees often change their jobs when they are not satisfied with it. Different factors may cause a change in the level of job satisfaction of an employee. For example, the dynamic nature of the Information Technology (IT) industry is an impactful factor that determines the job satisfaction of IT professionals. Foreseeing the employees' job satisfaction makes it easy for a company to take swift actions to improve the job satisfaction of its employees. In this research, we analyzed the effectiveness of machine learning (ML) methods for predicting job satisfaction using employee job profiles. There are job-specific factors in each job domain, and those factors may influence job satisfaction levels. Therefore, this research focused on the following fundamental questions: 1) How do existing ML models perform when predicting job satisfaction of software developers? 2) Can the job satisfaction prediction models be generalized to the other job roles in the IT industry? This study compared the performance of classification models: Random Forest (RF), Logistic Regression (LR), Support Vector Machine (SVM), and Neural Network (NN) in predicting the level of job satisfaction. Our experiments used two benchmark datasets: Stack Overflow developer survey and IBM HR analytics dataset. The experimental analysis shows that both employee-related factors and company-related factors contribute similarly to predicting job satisfaction. On average, the above ML models predict the job satisfaction of software developers with an accuracy of around 79%.**

Keywords - classification models, data mining, job satisfaction, machine learning

I. INTRODUCTION

Human resource is the most important factor for the success of any organization. Therefore, most organizations and companies are seeking talented, knowledgeable and experienced candidates for their job openings. Due to the technological advancements and complex lifestyles of the modern people, the current job market shows rapid changes. New jobs are being created, and some of the existing jobs have been taken over by new technologies. For example, robots are serving at some of the airports to do certain tasks which were performed by human employees. With these drastic changes, employees are migrating to demanding jobs to discover their passion and to satisfy their life expectations. Job satisfaction is an important aspect due to the fact that it represents an overall summary of how an individual feel about a lifetime of work [1]. Therefore, job satisfaction can be described as a pleasurable or positive emotional state from the appraisal in any field of interest. Employees who are satisfied with their jobs have the enthusiasm to drive the company towards success while improving themselves. Thus, employee job satisfaction is a vital factor that needs to be considered in the recruitment process. However, it is a challenging task to select the most suitable candidate from a plethora of applicants. There are popular filtering mechanisms used in human resource departments, which are mostly manual processes. For example, filtering candidates based on different factors in their resumes such as working experience and educational background. Owing to the new technologies and innovations, companies are moving towards novel techniques to make decisions regarding new recruits [2]. If the Human Resource (HR) managers can foresee the job satisfaction of a person, it will bring numerous benefits in terms of competitive advantage and efficiency in the recruitment process. On the other hand, it is beneficial for the employees to choose jobs with high job satisfaction. Different factors may influence the level of job satisfaction of an employee. For example, social, cultural and political factors such as employee salary, age, education level, and the complexity of the work to be done are some of the main influential factors for the level of job satisfaction. Nevertheless, the causes of employee job satisfaction or dissatisfaction mainly depend on the field that the employee works.

Various methods are available to predict job satisfaction. However, to the best of our knowledge, existing research is not focusing on predicting job satisfaction using machine learning (ML) techniques considering both employees' background data and company-related factors. In this research, we analyzed the performance of several ML models based on two case studies namely Stack Overflow developer surveys [12][26][27] and IBM HR analytics [25]. Different features extracted from the Stack Overflow developer survey were used to predict the job satisfaction of software developers. Then the study was extended to generalize the prediction model for predicting job satisfaction of other job roles in the IT industry (generalized model). Features extracted from IBM HR analytic dataset were used for the generalized model. We considered four different classifiers namely Random Forest (RF), Logistic Regression (LR), Support Vector Machine (SVM) and Neural Network (NN) for the prediction models. The objectives of this research are as follows:

- Studying the available approaches for predicting job satisfaction.
- Identifying the main influential factors of job satisfaction of the IT professionals.
- Exploring the possibility of generalizing job satisfaction prediction models to other job roles in the IT industry.

The rest of the paper is organized as follows. In the following section, we discuss a selected set of existing research studies related to our topic. In section 3, we present our empirical analysis of prediction models and the performance of each model. Then the findings are discussed with the results of the experimental analysis. Finally, we conclude this paper with the future directions of the research.

II. RELATED WORK

The advancement of internet technologies has allowed acquiring insights for accurate decision making through analytical formulas and data processing techniques. Employee related decision making such as employee turnover, attrition and job involvement is a crucial task as optimistic employees are the key success factors of a company. Therefore, recent researchers have focused on exploring the applicability of ML models in employee-related decision making [2]-[6]. Job satisfaction data mining has been widely used to extract meaningful knowledge about employee satisfaction. This approach is applicable for various domains and contexts in predicting the satisfaction level, identifying the most affecting factors of job satisfaction and taking remedial actions to improve the performance of employees [7] [8]. Even though both job satisfaction and career satisfaction are related to global life satisfaction, these two are independent of each other. For instance, while career satisfaction is related to turnover intention and leaving in the IT field, job satisfaction of IT professionals is highly related to employee turnover, which is a persistent problem in the IT industry [1]. It has shown that the level of job satisfaction strongly affects the turnover intention of software developers [3]. Employee job satisfaction is based on both objective and subjective data [9]. For example, a research study has been carried out to find the impact of family factors and the role of work in predicting career satisfaction. It was evaluated by collecting data from 344 participants through an online survey. In there, hypothesis testing has confirmed that there is a significant relationship between job satisfaction and work-family balance in improving the level of job satisfaction [10]. Many companies collect and keep employee records and data to study their job satisfaction. However, influential factors of job satisfaction may differ based on the industry, job role as well as the country and the region. For example, recent research has shown that personal development opportunities, relationship with the supervisor, and adherence to the duty roster are the most important factors for job satisfaction in the hospitality industry in the Alpine region [28].

A considerable number of research studies have been conducted using the data extracted from Stack Overflow as it is a world popular Q&A platform for software developers. However, the majority of them are related to the questions and answers [11],[12] posted in the Stack Overflow website rather than the Stack Overflow developer survey responses. Most of the existing research studies on predicting career/job satisfaction in different disciplines have used mostly statistical analysis methods rather than using sophisticated ML techniques. Therefore, it is worth exploring the potential of ML methods in predicting job satisfaction. ML is a branch of Artificial Intelligence (AI) that learns and improves automatically through experience. In there, classification is a supervised ML approach that uses label data to train the model which used to predict the labels of unknown examples. ML algorithms have been used for predicting, classifying and clustering various kinds of data in different domains and industries such as healthcare, financial and marketing. Most of the previous ML-based forecasting have been conducted as empirical analysis by comparing the performance of existing ML models. For example, a research study has examined the performance of the five existing classification algorithms when predicting the likelihood of hospital readmission [13]. According to their comparison, SVM has shown the best performance among the chosen algorithms, while the results of LR and Naïve Bayesian (NB) are lower than the other classifiers. Besides job satisfaction, satisfaction level prediction is another area of forecasting that has applied in different domains. For example, the customer satisfaction level prediction is used to improve products and services. Since companies are not only relying on product quality but even more on a service quality level, there is a significant need for identifying the customer satisfaction level. Thus, a research study of predicting customer dissatisfaction has been carried out using five existing classification models [14].

Ensemble ML algorithms such as RF are widely used for both classification and regression problems due to their excellent accuracy, ease of use and robustness. This is because the method of combining multiple independent learning algorithms increases the predictive performance that could be obtained from any of the single learners alone. To reduce the learning time and the computational cost, the fast algorithms such as decision trees are widely used in ensemble methods [15]. Binary classification is the most commonly used classification type where the target variable has only two classes. Researchers have shown that decision trees and NN perform well in binary classification through several studies [16] [17]. In addition, SVM, Decision Tree, RF and NB can be used as multi-class classifiers. For instance, student academic performance prediction using their academic progress, personal characteristics and behaviors relating to learning activities [18] [19] are two case studies which have used multi-class classification. Therefore, classification models are ideal for predicting the level of job satisfaction.

III. JOB SATISFACTION OF IT PROFESSIONALS

According to the recent analysis, healthcare and information technology(IT) related jobs are the top-rated jobs in the world. As a result, there has been tremendous growth in the software and IT industry over the last few years. Software development ranked as a top demanding job and software engineering has been rated as one of the rapidly expanding sectors in the world. Although the demand for software developers is nothing new, it has seen a significant rise in the last couple of years. According to the predictions, employment of software developers will increase by 22% from 2019 to 2029, which is much faster than the average of all other occupations [20]. Therefore, more employees are moving into the software development industry. However, IT related job specific factors may influence the level of job satisfaction of IT employees. For example, since software development is often a deadline-oriented process, the level of stress among software developers tends to be high. This is especially common among the less experienced developers. Moreover, adapting to rapidly changing cutting edge technologies is one of the most challenging tasks for software developers. Even though the idea of flourishing happiness among developers is often promoted by software companies, because of the above reasons, the IT industry has become the industry with the highest turnover rate in 2018 [21]. Therefore, the present work analyzed the factors which influence the job satisfaction of IT employees. This experimental analysis consists of three tasks which are shown in Fig. 1.

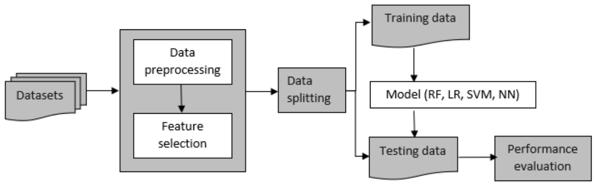


Fig. 1. Proposed methodology

In the first task, we retrieved data from the data sources and preprocessed to remove the noise. The second task was feature engineering and selecting the most discriminative features for training ML models. Finally, the prediction performances of the trained ML models were tested.

A. *Data*

The ever-increasing volumes of data and information shared on social media and collaborative sites have become a rich and valuable source of knowledge for a wide spectrum of research needs. When there is a need to learn about a new topic or to answer a particular query, people look for fast access to relevant information sources that would help them address that need. In the IT industry, software developers often visit online question and answering (Q&A) sites to find answers for their coding problems. Stack overflow is a well-known free Q&A website for IT professionals and enthusiastic software developers. Each year, Stack overflow collects data from the software developer community and makes the anonymized data available for researchers and other interested parties. This is named as the "Stack Overflow developer survey" which provides highly accurate data about software developers all around the world. Hence, we choose Stack Overflow developer survey datasets (dataset1) which have been released recently in three consecutive years: 2018, 2019 and 2020 [12] [26] [27]. The dataset1 was used for training the job satisfaction prediction model for software developers. It is mainly composed of categorical data such as Country, Developer Type, Gender, etc. Researchers use this public dataset for retrieving insights of the behavior of IT employees [22]. In 2018, they published their Annual Developer Survey results for the eighth consecutive year with the largest number of respondents yet [23]. Responses have been collected in January 2018 and nearly 100,000 developers have responded to this 30-minutes survey. Apart from the Stack overflow dataset, International Business Machines (IBM) HR analytic dataset (dataset2) [25] was used to analyze job satisfaction of both IT and non-IT employees in the IT industry. This dataset consists of job-related features common for employees in many industries such as age, job role, monthly income, education, etc. The volume/size and the number of features of each dataset before the preprocessing stage are shown in Table I.

TABLE I. COMPOSITION OF EACH DATASET

Dataset	Size	Features
StackOverflow developer survey 2018	98,855	129
StackOverflow developer survey 2019	88,883	85
StackOverflow developer survey 2020	64,461	61
IBM HR analytic	1,500	35

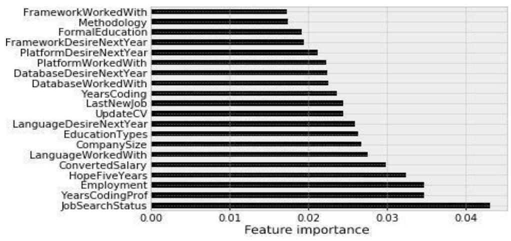
B. *Experimental design*

When datasets become bigger in both volume and variety with a large number of features, it is necessary to apply ML techniques to extract patterns and knowledge from the data. Effective data preprocessing and feature engineering techniques are vital for better performance of ML models. Hence, this study used two-dimensionality reduction techniques to select features from the dataset. First, unique identifiers such as "response_id" were removed from datasets as they do not hold any significant importance to the analysis. Then the features which have more than 50% of missing values were removed. Considering the RF feature importance, 53 features were selected from dataset1 to train ML models for predicting the job satisfaction of software developers. Only 33 features were considered among 35 features in dataset2 to train the generalized model to predict job satisfaction of other job roles in the IT industry. Since the majority of the selected features were categorical, missing values in the selected features were replaced with the mode. Even though the chosen ML algorithms are robust to the over-fitting problem, we removed the classes with fewer frequencies in some features such as gender. For example, we considered the users whose gender is either male or female and removed the other gender categories which have very few examples in the dataset1. Since most of the ML algorithms accept only numerical data, categorical data were converted into numerical values using Label encoding. The label is job satisfaction in both scenarios. It has seven classes in the dataset1 namely, Extremely Satisfied, Moderately satisfied, Slightly satisfied, Neither satisfied nor dissatisfied, Slightly dissatisfied, Moderately dissatisfied and Extremely dissatisfied. However, we made it into three classes for better performance by grouping the first three classes into one class called 'Satisfied' and the last three classes into one class called 'Dissatisfied' and remaining the class 'Neither satisfied nor dissatisfied' as it is. The label has four classes in the dataset2, but were made into three classes. After the preprocessing and feature engineering stages, datasets were split into training and testing data such as 80% of the dataset for training the classifiers and 20% for testing.

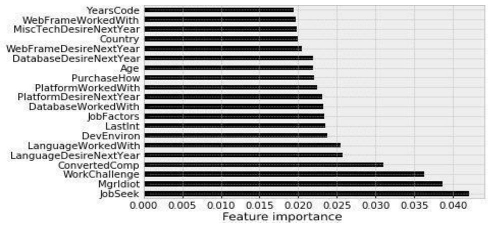
In this research, we used supervised classification methods. Since the output variable has more than one class in both scenarios, the multi-class classification technique was used to classify job satisfaction. We compared four different predictive models namely, Random Forest (RF), Logistic Regression (LR), Support Vector Machine (SVM), and Neural Network (NN) to see the difference of the performance in predicting job satisfaction of software developers. These four algorithms have been selected due to their flexibility in handling a range of classification problems with a large feature space [15].

IV. RESULTS AND DISCUSSION

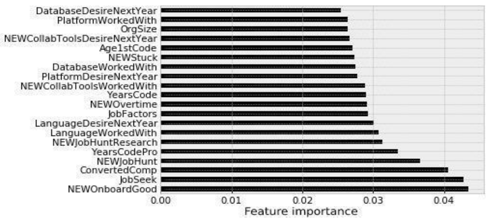
This section discusses the results of the experiments and the limitations of the study with future directions. After removing the noisy data in the preprocessing stage, we considered 97,869 records, 87,740 records and 63761 records for this study from Stack Overflow developer survey 2018, 2019 and 2020 respectively. The total number of 1472 records were considered from the IBM HR analytic dataset to train and test the generalized model for predicting job satisfaction of both IT and non-IT employees in the IT industry. The variable or the feature importance provides the statistical significance of the variables in the dataset. This is very important when using the multi-class classification methods to make predictions as it can be used to identify whether the selected features contribute or do nothing in classification with the chosen ML models. In this experiment, a total number of 53 features were selected as the most important features from Stack Overflow developer survey datasets for predicting the job satisfaction of software developers. A total number of 33 features were selected as the most discriminate features for predicting the job satisfaction of other job roles in the IT industry



(a) 2018



(b) 2019



(c) 2020

Fig. 2. Feature importance of Stack

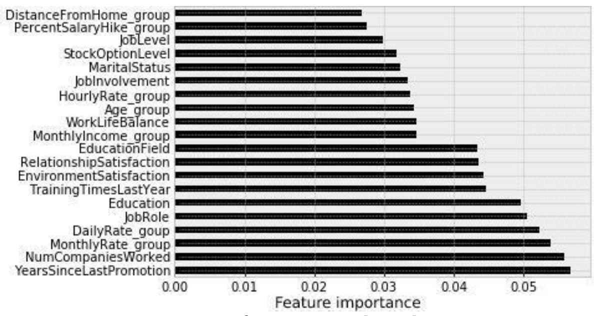


Fig. 3. Feature importance of IBM HR analytic dataset

According to the descriptive statistical analysis and the graphs of feature importance (figure 2 & figure 3), some variables in the dataset are less significant than some other variables for predicting the level of job satisfaction. For example, it shows that the contribution of the feature, "jobSeek" is one of the most significant features. In addition, graphs shown in Fig. 2 show the variations of job satisfaction influential factors for software developers in past consecutive years. Overall, common main influential factors for deciding the level of job satisfaction of software developers are as follows:

- Availability of training and managerial support
- Monthly income
- Years of coding experience
- Company size
- Challenges in workplace
- Programming languages work with
- Platforms work with

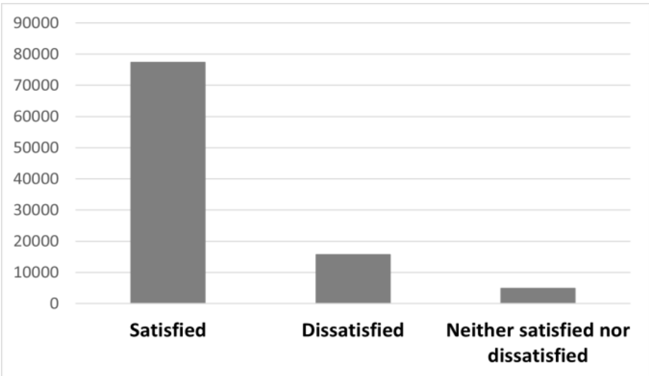
Following are the key factors to decide the level of job satisfaction of both IT and non-IT job roles in the IT industry as shown in figure 3.

- Promotions
- Number of companies worked with
- Monthly income
- Job role
- Education
- Training
- Environmental satisfaction
- Relationship satisfaction

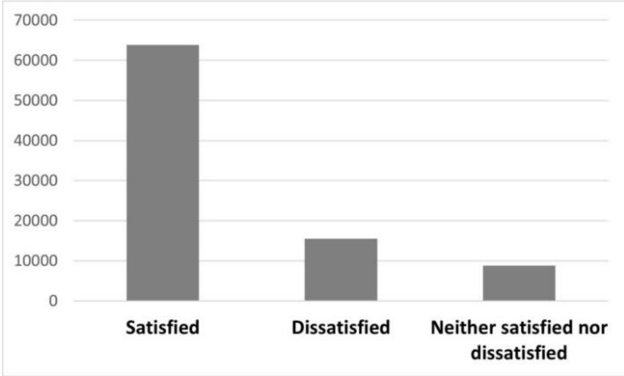
Feature important graphs shown in figure 2 & figure 3 show that both employee-related factors and company-related factors are contributing similarly when deciding the level of job satisfaction of IT employees. For example, providing training opportunities, monthly income, promotions and company environment are a few of the factors that HR managers and companies could directly involve to seed a high level of job satisfaction among employees

The problem of class imbalances causes a decrease in the accuracy of the predictive models. There are minority classes in the labels of all datasets. Thus, we reduced the classes into three by aggregating similar classes. Although the reduction of the number of classes increased the data in each class, the class imbalance still presents as shown in Fig. 4. Therefore, Synthetic Minority Oversampling Technique (SMOTE) [24] was used to synthesize new examples for the minority classes. After reducing the number of classes and removing the class imbalance, the accuracy of each model increased. For example, RF model performance comparison with 7 classes with the class imbalance, and with 3 classes after applying SMOTE is shown in Table II. Accuracy is not a good indicator of model performance in this study due to class imbalances. Because it is biased as the rare classes can be masked by the majority classes. Thus, we used four performance measures namely accuracy, precision, recall and f1-score as the evaluation criteria for this study. Moreover, hyperparameter tuning was used to improve the performance of each model. For instance, hyperparameters in RF are (1) maximum depth: the maximum depth of the tree (2) maximum features: the maximum number of features Random Forest is allowed to try in an individual tree and (3) number of estimators: the number of trees in the forest. A grid search was performed over the specified parameter values using the cross-validation technique to assess model performance and to find the best set of parameters. The best parameter value of maximum depth is 12, maximum features are 50, and the number of estimators is 25 for RF. Then the SVM hyperparameters were tuned with Radial Basis Function (RBF) kernel function and the best parameter value for both C and gamma is 1.

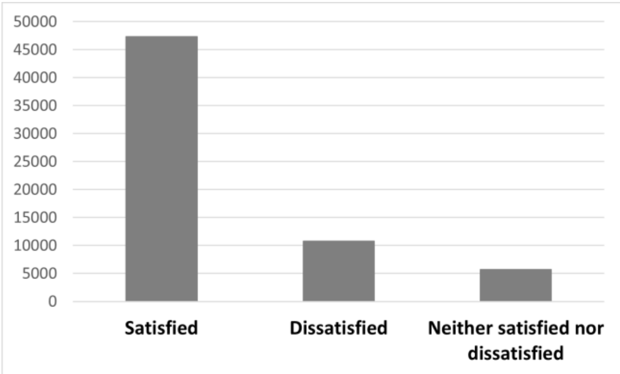
A feed-forward NN model was built using Keras and TensorFlow. We created a fully connected network with two hidden layers. Because of the advantages of computational efficiency and non-linearity, we used the "relu" activation function for the input layer and the hidden layers. Since this NN model is for multi-class classification, the "softmax" activation function is used for the output layer. Finally, the network used the efficient Adam gradient descent optimization algorithm and logarithmic loss function, "sparse_categorical_crossentropy" for compilation. With these parameters, RF shows the highest accuracy, precision, recall and f1-score among the chosen classifiers for all the datasets when predicting job satisfaction of software developers.



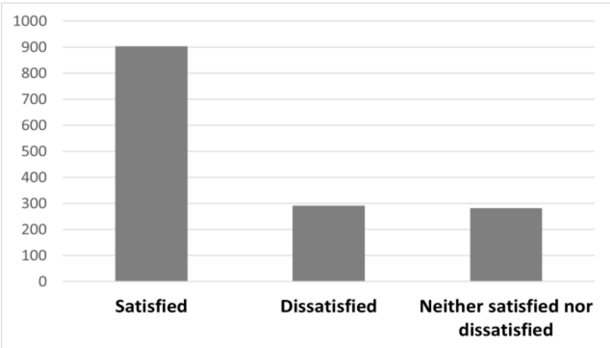
(a) 2018



(b) 2019



(c) 2020



(d) 2020 IBM HR analytic

Fig. 4. Class distribution of labels in Stack Overflow developer survey (a)2018, (b)2019 & (c)2020 and (d)IBM HR analytic dataset

TABLE II. RF MODEL PERFORMANCE WITH 7 CLASS LABEL VS 3 CLASS LABEL

	Label with 7 classes				**Label with 3 classes**			
	Accuracy	***Precision***	***Recall***	***f1-score***	***Accuracy***	***Precision***	***Recall***	***f1-score***
RF	0.68	0.62	0.68	0.62	0.80	0.74	0.80	0.75

This is because RF is an ensemble algorithm that consists of a group of decision trees. Table III shows that the RF shows 80% accuracy of job satisfaction of software developers while others show around 79% accuracy.

TABLE III. EVALUATION METRICS OF CLASSIFICATION MODELS

Dataset	ML model	Accuracy	Precision	Recall	f1-score
Stack Overflow 2018	RF	**0.80**	**0.74**	**0.80**	**0.75**
	SVM	0.79	0.62	0.79	0.69
	LR	0.79	0.69	0.79	0.70
	NN	0.79	0.62	0.79	0.69
Stack Overflow 2019	RF	**0.76**	**0.68**	**0.76**	**0.71**
	SVM	0.73	0.53	0.73	0.62
	LR	0.74	0.68	0.74	0.67
	NN	0.73	0.53	0.73	0.61
Stack Overflow 2020	RF	**0.76**	**0.69**	**0.76**	**0.70**
	SVM	0.74	0.55	0.74	0.63
	LR	0.75	0.65	0.75	0.67
	NN	0.74	0.55	0.74	0.63
IBM HR analytic	RF	*0.33*	*0.31*	*0.33*	*0.31*
	SVM	*0.38*	*0.33*	*0.38*	*0.28*
	LR	*0.37*	*0.35*	*0.37*	*0.34*
	NN	*0.35*	*0.35*	*0.35*	*0.35*

According to the results in the latter section in table III, it shows that the classifiers RF, SVM, LR and NN are not performing well with the IBM HR analytic dataset. Therefore, the above classifiers cannot predict employees' job satisfaction as a generalized model for predicting job satisfaction of other jobs in the IT industry. These results reveal that job-specific factors have a high contribution in deciding the level of job satisfaction. For example, above mentioned models performed well with predicting job satisfaction of software developers, and years of coding experience, programming languages work with and platforms work with are some of the most significant job-specific factors apart from the salary, training and workplace challenges. These factors are based on a globally collected dataset, and this experiment can be further extended to see the applicability of the above factors for the local IT industry by collecting local datasets.

The present study is a first step towards forecasting the level of job satisfaction of software developers using ML models. However, this can influence the software developer's survival in the software industry and further aid in the recruitment process. The findings of this study help HR managers to improve the identified company-related factors which caused a change in the level of job satisfaction of employees. It will increase the company reputation directly and indirectly through positive behavior and commitment of employees. For example, establishing training programs for newly recruited employees, giving promotions and career development opportunities, and building good relationships are worth the success of a company through highly satisfied employees. In addition, the results of this study are beneficial for software developers to choose job opportunities where they can gain high job satisfaction considering their background data and company-related factors. As future works, three directions can be followed as follows. (1) The accuracy of the model can be increased by including more training data which is collected from different sources other than the Stack Overflow developer survey. (2) This work can be extended by implementing a NN model that finds the best weights of each factor for job satisfaction. (3) The effect of the "work from home" approach can be analyzed to change the job satisfaction level among employees in the IT industry.

V. CONCLUSION

In this research, we investigated supervised ML models for predicting the job satisfaction of IT employees. Prediction performance of multi-class classifiers namely RF, SVM, LR and NN were compared using two benchmark datasets. Accuracy, precision, recall and f1-score were used as the performance metrics to evaluate and compare the classifiers. The experimental results show that the above ML models can predict the level of job satisfaction of software developers using their background data and company-related data with an accuracy of around 79%. Further, we investigated the performance of the aforementioned classifiers when predicting job satisfaction of both IT and non-IT employees in the IT industry. It reveals that the above classifiers cannot be utilized as generalized models to predict the job satisfaction of IT-related employees who are not software developers. In addition, seven (07) factors were identified as the most influential factors of job satisfaction of software developers. In summary, the findings of this study are beneficial for several parties such as IT employees, IT related companies and researchers in this domain.

REFERENCES

[1] J. Lounsbury, L. Moffitt, L. Gibson, A. Drost, and M. Stevens, “An investigation of personality traits in relation to job and career satisfaction of information technology professionals,” JIT, vol. 22, pp. 174–183, 03 2007.

[2] F. Fallucchi, M. Coladangelo, R. Giuliano, and E. William De Luca, “Predicting employee attrition using machine learning techniques”, vol. 9, no. 4, 2020.

[3] V. Wickramasinghe, “Impact of time demands of work on job satisfaction and turnover intention: Software developers in offshore outsourced software development firms in sri lanka,” Strategic Outsourcing: An International Journal, vol. 3, pp. 246–255, 11 2010.

[4] P. Rohit and P. Ajit, “Prediction of employee turnover in organizations using machine learning algorithms,” International Journal of Advanced Research in Artificial Intelligence, vol. 5, 10 2016.

[5] Y. Choi and J. Choi, “A study of job involvement prediction using machine learning technique,” International Journal of Organizational Analysis, vol. ahead-of-print, 08 2020.

[6] A. A. A. Khaled Alshehhi, Safeya Bin Zawbaa and M. U. Tariq, “Employee retention prediction in corporate organizations using machine learning methods,” Academy of Entrepreneurship Journal, vol. 27, 08 2021.

[7] M. Murawski, N. Payakachat, and C. Koh-Knox, “Factors affecting job and career satisfaction among community pharmacists: A structural equation modeling approach,” Journal of the American Pharmacists Association: JAPhA, vol. 48, pp. 610–20, 09 2008.

[8] A. Domagala, J.-N. Pena-SA˜ nchez, and K. Dubas-Jakobczyk, “Career satisfaction of polish physicians - evidence from a survey study,” European Journal of Public Health, vol. 29, 11 2019.

[9] A. Altamimi, "Literature on the relationships between organizational performance and employee job satisfaction," Archives of Business Research, vol. 7, 2019.

[10] N. Gopalan and M. Pattusamy, "Role of work and family factors in predicting career satisfaction and life success," International Journal of Environmental Research and Public Health, vol. 17, p. 5096, 07 2020.

[11] S. Wang, D. Lo, and L. Jiang, "An empirical study on developer interactions in stackoverflow," 03 2013, pp. 1019–1024.

[12] A. Joorabchi, M. English, and A. Mahdi, "Text mining stackoverflow: Towards an insight into challenges and subject-related difficulties faced by computer science learners," Journal of Enterprise Information Management, vol. 29, pp. 255–275, 03 2016.

[13] S. Alajmani and H. Elazhary, "Hospital readmission prediction using machine learning techniques," International Journal of Advanced Computer Science and Applications, vol. 10, 01 2019.

[14] S. Meinzer, A. Thamm, U. Jensen, J. Hornegger, and B. Eskofier, "Can machine learning techniques predict customer dissatisfaction? a feasibility study for the automotive industry," Journal of Artificial Intelligence Research, vol. 6, pp. 80–90, 01 2017.

[15] S. Ahamed and E. Daub, "Machine learning approach to earthquake rupture dynamics," 06 2019.

[16] Y. Alejandro and L. Palafox, Gentrification Prediction Using Machine Learning, 10 2019, pp. 187–199.

[17] G. Deepali, A. Brar, and P. Sandhu, "Modeling of fault prediction using machine learning techniques," 08 2020.

[18] Thi, H. Dinh, T. Pham, C. Loan, G. Nguyen, N. Thi, and N. Thi Lien Huong, "An empirical study for student academic performance prediction using machine learning techniques," International Journal of Computer Science and Information Security, vol. 18, 04 2020.

[19] M. Asim and Z. Khan, "Mobile price class prediction using machine learning techniques," International Journal of Computer Applications, vol. 179, pp. 6–11, 03 2018.

[20] Bureau of Labor Statistics, U.S. Department of Labor, Occupational Outlook Handbook, Software Developers, 2019 (Accessed on: June 13, 2021). [Online]. Available: https://www.bls.gov/ooh/computer-and-information-technology/software-developers.htm

[21] P. Petrone, See The Industries With the Highest Turnover (And Why It's So High), 2018 (Accessed on: August 02, 2020). [Online]. Available:https://www.linkedin.com/business/learning/blog/learner-engagement/see-the-industries-with-the-highest-turnover-and-why-it-s-so-hi

[22] T. Ahmed and A. Srivastava, "Understanding and evaluating the behavior of technical users. a study of developer interaction at stackoverflow," Human-centric Computing and Information Sciences, vol. 7, 12 2017.

[23] StackOverflow, "Stack overflow 2018 developer survey," 2018 (Accessed on: January 27, 2021). [Online]. Available: https://www.kaggle.com/stackoverflow/stack-overflow-2018-developer-survey

[24] S. Uyun and E. Sulistyowati, "Feature selection for multiple water quality status: Integrated bootstrapping and smote approach in imbalance classes," International Journal of Electrical and Computer Engineering, vol. 10, pp. 4331–4339, 08 2020.

[25] pavansubhash, "IBM HR Analytics Employee Attrition & Performance," *Kaggle.com*, 2017. https://www.kaggle.com/pavansubhasht/ibm-hr-analytics-attrition-dataset.

[26] M. Chirico, "Stack Overflow Developer Survey Results 2019," *kaggle.com*, 2019. https://www.kaggle.com/mchirico/stack-overflow-developer-survey-results-2019?select=survey_results_public.csv (accessed Aug. 22, 2021).

[27] A. Khan, "Stack Overflow Developer Survey 2020," *www.kaggle.com*, 2020. https://www.kaggle.com/aitzaz/stack-overflow-developer-survey-2020 (accessed Aug. 22, 2021).

[28] P. Heimerl, M. Haid, L. Benedikt, and U. Scholl-Grissemann, "Factors Influencing Job Satisfaction in Hospitality Industry," SAGE Open, vol. 10, no. 4, p. 215824402098299, Oct. 2020.