Bachelor Thesis

"What Influences Employee Job Satisfaction the Most: Machine Learning Approach"

Author Boriss Verbickis, BITL1

Supervisor Māra Pudāne, M.sc.ing

Riga 2023

Executive Summary

A business's success largely depends on employee satisfaction, significantly affecting workers' productivity, motivation, and ability to achieve its goals. A company needs to identify the factors influencing employee job satisfaction to develop effective strategies to retain critical top-performing employees and improve overall employees productivity. This study aims to identify the factors that significantly impact employee job satisfaction using a machine learning approach based on the collected data using Job Satisfaction Survey by Paul Spector. Understanding the key factors influencing job satisfaction will allow comparing different machine learning algorithms' performance and accuracy for the set task by evaluating their performance metrics. In the research process, a survey was conducted to collect data from 128 respondents from three offices in one specific organization on nine factors affecting their job satisfaction. A machine learning data analysis approach was then applied to determine the impact of various factors affecting job satisfaction. Specifically, the author used three machine learning algorithms: Random Forest, XGBoost, and Linear Regression. Additionally, the author used feature importance techniques to identify the most significant variables that contribute to the model's predictive accuracy: default feature importance, permutation feature importance, and drop-column feature importance. To get a more complete and accurate result that affects job satisfaction the most, several machine learning algorithms and methods were used in the study. The study results showed that the nature of work and promotion significantly affect employee job satisfaction. On the other hand, pay had a relatively small effect on employee job satisfaction. Additionally, there were variations in the predictive accuracy of different algorithms. The Random Forest algorithm performed the best among the other algorithms, with

the lowest MSE and highest R^2 values. Meanwhile, the XGBoost algorithm had the highest MSE and lowest R^2 values, indicating that it was the least accurate algorithm for predicting job satisfaction. It is important to note that the optimal algorithm may vary depending on the research problem. Therefore, several algorithms should be utilized and tested to maximize redundancy. The resulting benefits are significant for organizations seeking to improve employee job satisfaction and retention rates. By identifying the key factors influencing job satisfaction, organizations can develop effective strategies to improve them and create a more positive work environment for their employees.

The study has some limitations that should be noted. First, the data were collected from a limited sample size that may accurately reflect only a portion of the population. Second, the study was conducted in a specific industry, which may not apply to other industries. Finally, the study did not consider individual differences in job satisfaction, such as personal characteristics and values. In terms of future work possibilities, the author recommends replicating the study in different sectors and with a larger sample size to improve the applicability of the results across different sectors. Furthermore, future research could examine the influence of individual differences, such as personality traits, cognitive abilities, physical characteristics, and cultural backgrounds, on job satisfaction and the effectiveness of different strategies in improving satisfaction. Overall, this research provides valuable insight into the application of machine learning approaches to determine what influences job satisfaction, which can benefit organizations by improving employee retention, motivation, and productivity.

The bachelor thesis is written in English and has 59 pages, 13 figures, three tables, and 41 literature sources.

Keywords: job satisfaction, predictors, machine learning applications, feature importance methods

Abstract

Uzņēmuma efektivitāte lielā mērā ir atkarīga no darbinieku apmierinātības līmeņa, kas ietekmē viņu produktivitāti, motivāciju un spēju sasniegt uzņēmuma mērķus. Uzņēmumam ir svarīgi identificēt faktorus, kas ietekmē darbinieku apmierinātību ar darbu, vērtīgo darbinieku noturēšanai un vispārējas darbinieku produktivitātes uzlabošanai. Pētījuma mērķis ir noteikt faktorus, kuriem ir vislielākā ietekme uz darbinieku apmierinātību ar darbu, izmantojot mašīnmācīšanās pieeju. Dati tika apkopoti, aptaujājot 128 respondentus un izskatot deviņus faktorus, kas varētu ietekmēt viņu apmierinātību ar darbu. Datu analīzei tika izmantoti trīs mašīnmācīšanās algoritmi un trīs faktoru svarīguma noteikšanas metodes. Rezultāti liecina, ka "Darba Raksturs" un "Paaugstinājums Amatā" faktoriem ir vislielākā ietekme uz darbinieku apmierinātību ar darbu, savukārt atalgojumam tika konstatēta salīdzinoši neliela nozīme.

Darba rezultāti ir noderīgi uzņēmumiem, kuru mērķis ir labāko darbinieku noturēšana un kopējās produktivitātes uzlabošana. Galveno darba apmierinātības faktoru noteikšana palīdzēs stratēģiju izstrādē labākas darba vides radīšanai. Pētījumam piemīt daži ierobežojumi. Datu iegūšanai, tika aptaujāta neliela cilvēku grupa, tādēļ rezultāti var būt nereprezentatīvi visiem gadījumiem. Pētījums tika vērsts uz konkrētu nozari, tāpēc rezultāti var nebūt vispārināmi citām sfērām. Pētījumā tika izmantoti tikai trīs mašīnmācīšanās algoritmi un trīs iezīmju svarīguma noteikšanas metodes, un netika izmēģināti citi algoritmi vai metodes, kas varētu radīt atšķirīgus rezultātus. Turpmākajiem pētījumiem būtu lietderīgi atkārtot pētījumu citās nozarēs un ar lielāku aptaujāto darbinieku skaitu, lai palielinātu rezultātu precizitāti un izmantošanas iespēju ārpus darbā pētītās uzņēmuma nozares. Ieteicams izmantot papildus algoritmus un metodes arī lai nodrošinātu, ka to atklājumi ir precīzi un vispārināmi. Pētījuma rezultāti atklāj, kā

mašinmācīšanās izmantošana ir iespējama darba apmierinātības ietekmējošo faktoru identificēšanai. Ņemot vērā rezultātus, uzņēmumi varēs uzlabot darba apstākļus, paaugstināt darbinieku motivāciju un produktivitāti.

Diplomdarbs ir uzrakstīts angļu valodā, tā apjoms ir 58 lapaspuses, tajā ir iekļautas 13 figūras, trīs tabulas, un 41 literatūras avoti.

Atslēgvārdi: apmierinātība ar darbu,nosakošie faktori,mašīnmācīšanāslietojumi, iezīmju nozīme

Table of Contents

Introduction

Importance of Topic

Job satisfaction is one of the critical topics in Human Resource Management and Psychology. Organizations have become more competitive and performance-oriented, which results in retaining critical top-performing individuals. For an organization to keep high-performing employees, the entity must ensure a high satisfaction rate (Rehman et al., 2013). Job satisfaction is a critical determinant of employee retention, productivity, motivation, and commitment to organizational goals (Spector, 1985). Employees with higher satisfaction rates are likelier to stay with the organization and less likely to exhibit counterproductive work behavior, absenteeism, and turnover intentions (Judge et al., 2001).

However, keeping employees' job satisfaction at an elevated level is challenging. Some organizations underestimate the importance of this psychological attitude, while other companies need more resources to keep the rate high (Harper et al., 2015). In Psychology, there is already a set of methods that allow to determine employee satisfaction, such as interviews or surveys, containing many questions. However, there might be a better solution for a digital age, namely reducing the number of factors which would then enable, for example, investment in automatic detection of the factors affecting job satisfaction, which would allow reduce the number of questions asked to people. According to Harper et al. (2015), even successful organizations should identify the factors related to job satisfaction as they affect performance and retention. Identifying the factors contributing to job satisfaction is crucial for organizations to develop effective strategies to keep valuable staff members and enhance their

overall performance. Therefore, it is essential to determine the most critical factors that affect job satisfaction.

Statistical methods and machine learning algorithms that learn from previous experience can help solve such problems and identify the most vulnerable parts of the organization related to job satisfaction (Y. Chen et al., 2022). The machine learning approach might be a complex solution compared to non-IT existing methods, which include surveying employees and conducting interviews and focus groups. Nevertheless, if trained well, the results would be more accurate and precise than the findings by a human expert (De Togni et al., 2021). By identifying the most critical parts of the organization, the employer will be aware of the components that must be improved in the first place to make employees more satisfied. This research aims to identify the factors that stand apart in their importance over job satisfaction and influence it the most, resulting in identifying which key variables should be focused upon to maximize the improvement of job satisfaction overall.

To solve the problem, the research adopts a machine learning approach that utilizes three different algorithms, including Random Forest, XGBoost, and Linear Regression, along with three feature importance methods: default feature importance, permutation feature importance, and drop-column feature importance. The author utilized Paul Spector's Job Satisfaction Survey (JSS), which consists of questions about several attributes of job satisfaction. The author collected 128 responses, which were then processed by machine learning algorithms to identify patterns and relationships between the survey questions and overall job satisfaction. Data were split into training and testing data sets during the data analysis to evaluate the quality and performance of the algorithms used in the research. Further, the Mean

Squared Error (MSE) and R Squared (R^2) scores were calculated for each algorithm on the testing data sets.

The results of this study revealed that across all three algorithms and three feature importance methods, "Nature of Work" and "Promotion" were consistently identified as the two most significant predictors of job satisfaction. Even though the MSE scores were relatively high and R^2 scores were relatively low, the research provides valuable insights into the determinants of job satisfaction, particularly the importance of variables, using machine learning algorithms.

The research consists of five chapters. Chapter one introduces the topic and outlines the purpose and motivation of the research. Chapter two comprehensively reviews the literature on employee job satisfaction, its influencing factors, and machine learning-related concepts, namely machine learning frameworks and feature importance methods. Chapter three presents the survey design, data collection, and procedures for the research. Chapter four reveals the study results and findings in detail. Finally, chapter five discusses and concludes the results, their implications for organizations, research limitations, and future possibilities. Additionally, the appendix shows the survey and the link to GitHub, where the complete Python code used in the research is located.

Theory

Employee Job Satisfaction and Its Place in the Organization

According to Valentine et al. (2020), job satisfaction is an individual's positive feelings and evaluations of their employment. Cumbey and Alexander's (1998) analysis shows that job satisfaction is an exhilarating emotion that depends on the contact between employees, their qualities, beliefs, expectations, and the organization's environment. McNeese-Smith (1996) defines it as a worker's general attitude regarding their work. Knoop (1995) described it as an employee's general attitude toward work or specific aspects of work.

Job satisfaction is one of the most studied job attitudes in Psychology and Human Resource Management, which has been researched since the 1930s (Wright, 2015). Since employers want to reduce employee turnover and award critical top-performing workers, job satisfaction has become a frequently studied variable in organizational research (Spielberger, 2004). Employees use job satisfaction to measure their satisfaction with their current position and whether their work is fulfilling. Organizations use such terms to measure employees' performance and the likelihood of staying with the company. To motivate employees and encourage them to perform better, they must be satisfied at their jobs (Raziq & Maulabakhsh, 2015). Consequently, for organizations to remain competitive and efficient, employers should keep their employees fulfilled and satisfied.

Predictors of Job Satisfaction

Factors that influence job satisfaction must be determined first before determining their impact. The factors differ from research to research; however, for more detailed results, the author has chosen Paul Spector's Job Satisfaction Survey (JSS), which covers nine predictors

(Spector, 1985). The more factors covered in the research, the more specific answer will be in describing which of the predictors has the most significant effect on job satisfaction. According to Spector (1985), the Job Satisfaction Survey assesses nine dimensions linked to overall satisfaction with an average correlation coefficient of 0.70. The dimensions will be used as predictors in further study. Paul Spector's survey does not cover new technological trends such as hybrid working mode; nevertheless, all the predictors covered in the survey are relevant in modern organizations.

Pay. Every potential employee considers pay the main criterion when applying for a job. In the modern world, employees should be offered competitive pay packages compared to their outside competitors in the same field. Satisfaction comes from achieving fair and reasonable rewards (Sageer et al., 2012).

However, according to Qasim et al. (2012), a pay increase alone cannot improve job satisfaction. Job satisfaction cannot be determined solely by pay because even highly paid employees may remain unsatisfied if their job is not fulfilling. Pay is significant without any doubt; nevertheless, other values may be more vital to employees. Therefore, the author also considers other factors influencing overall work satisfaction in this research.

Promotion. Sageer et al. (2012) state that promotion can be viewed as a significant accomplishment that leads employees to slightly higher pay, responsibilities, authority, sovereignty, and prestige. As a result, the possibility for promotion and promotion itself is a significant element in determining employee job satisfaction. Previous research has found a close correlation between promotion opportunities and job satisfaction (Pergamit & Veum,

1999; Peterson et al., 2003). Therefore, employers should be aware of such motivation tools to ensure employees are willing to grow and perform at the top level.

Supervision. Effective communication and relationships with the supervisor are fundamental for employees to grow and become more competent. Constructive feedback, criticism, and appraisal provided by experienced specialists are crucial aspects that give space for improvement and self-development. Moreover, employees treated equally are more likely to be satisfied with their supervisors and, subsequently, have higher work satisfaction levels (Sageer et al., 2012).

Fringe Benefits. Fringe benefits, or non-monetary rewards, have historically been advocated as a practical method for organizations to boost their workers' total return on investment without raising salaries or providing other monetary benefits (den Dulk et al., 2012). Fringe benefits are essential elements regularly included in hiring packages. Such benefits include health insurance, life insurance, tuition aid, childcare reimbursement, employee stock plans, and personal use of a corporate vehicle.

According to Galanaki (2013), one of the main advantages for organizations to provide fringe benefits to employees is reaching economies of scale effects. Due to the large size of ordered services, organizations receive special offers with lower prices. At the same time, employees benefit from the value of the service, which would cost more if purchased individually.

Contingent Rewards. Valentine et al. (2020) state contingent rewards are performance-based compensations based on whether employees have met specific goals or how they performed. This method aims to pro the top-performing employees with more significant pay increases. The main reasons for having such increases are to reward and retain the best employees. In turn, employees will be pleased that the organization appreciates their work and results and will be more satisfied with their jobs.

At first glance, monetary rewards attract individuals more than other rewards. While this still can be true for some persons; nevertheless, research shows that contingent rewards such as work flexibility, public recognition, time off, and one-on-one meetings have a more substantial attractive power on the potential employee when applying for a position (Aguinis et al., 2013; Schlechter et al., 2015).

Operating Procedures. Danish and Usman (2010) state that rules, regulations, procedures, and job obligations are referred to operating procedures. Additionally, it includes the nature of the job and the organization's values to which one should adhere. Organizing procedures describe how employees must perform their duties in an organization. Especially for newcomers, following pre-written rules and procedures is more straightforward than inventing their approach to accomplish their tasks. Thus, if employees know their workplace obligations, they are more expected to be fulfilled with the job.

Co-workers. In organizations, colleagues are an integral part of the working life of employees. Projects and group work require a team of employees to work together to accomplish the task. Workplace incivility negatively impacts job satisfaction, while emotional exhaustion completely mediates relationships (Rhee et al., 2017). In addition, miscommunication and disagreement between co-workers can damage social relationships, negatively influencing job satisfaction and performance. Cooperation and harmony are practical approaches to improving employee satisfaction (Huang & Liu, 2017). A friendly co-worker relationship is always considered a good and efficient practice since it positively affects the working environment.

Nature of Work. According to Lerouge et al. (2014), the nature of work can be defined as the type of job an individual does. The nature of a job consists of attributes such as the purpose of the job, liking the job itself, enjoyableness of the job, skills required for the position, responsibilities at the workplace, and type of work that an individual does. Working at a job only since it is the source of obtaining money inevitably leads to burnout and, as a result, increases turnover and decreases job satisfaction. However, the question of personality yet remains. Some might hate specific jobs, but at the same time, there will be people who do not mind doing undesired work responsibilities.

Communication. Sageer et al. (2012) wrote that employees still need to discover many organizations' missions, visions, and goals. To maximize employees' talents and competencies, the employer can create a corporate culture that makes workers part of the organization. Employers should inform their employees about company positions, progress made, and how they contribute directly to the company's success. Building trust and accountability, explaining clear expectations, and securing employees' commitment are the cornerstones of an organization's success. Therefore, by building effective internal communication, an organization will become stronger and more competitive, while the employees of such an organization will be more satisfied.

The Importance of Machine Learning in the Current Era

During the last centuries, people have tried to simplify everyday tasks and processes by inventing new tools. Because of this reason, computers were developed. Nowadays, humanity cannot imagine life without computers and phones since such technologies make life easier, such as traveling, computing, and communication. Machine learning (ML) was developed to create reasoning like humans and to solve otherwise unsolvable problems. A machine learning algorithm is a mathematical model that computers use to learn how to accomplish specific tasks without explicit programming (Mahesh, 2019). Zhou (2021) describes machine learning as a technique for improving system performance using computational methods to learn from experience.

In his book, Zhou (2021) states that computers can learn from experience like humans. The difference between humans and machines is that experience is explicitly viewed as data in computer systems, and machine learning consists of developing algorithms that create models

from that data. The term model refers to the output learned from data. However, large data sets are required for machine learning algorithms to identify the regularity between variables and produce outcomes. The more data is provided to the algorithm, the more accurate the model is.

With the help of statistical methods, it becomes possible to train algorithms for classifications and predictions (IBM Cloud Education, 2022). By providing machine learning algorithms with considerable qualitative data, computers can make accurate predictions and classifications in almost every area, outperforming human specialists (De Togni et al., 2021). A wide range of fields in computer science is now utilizing machine learning, such as multimedia, graphics, network communication, software engineering, chipset design, and even architecture (Zhou, 2021). Therefore, the conclusion is that machine learning algorithms may also identify the factors that have the most significant effect on job satisfaction.

Machine Learning Algorithms

Machine learning algorithms vary in structure and are chosen depending on the goal and the training data size (Zhou, 2021). To get more accurate results, the machine learning model requires more data. However, sometimes the data size might be limited due to the specific area in which the data is collected. Therefore, the best approach would be to pick several algorithms with distinct characteristics to solve the same problem.

The data obtained from the Job Satisfaction Survey will then be used to train and test machine learning algorithms to forecast the importance of the factors with the most significant effect on overall job satisfaction. The survey uses a Likert scale to measure the degree of agreement or disagreement with a question or statement from 1 (strongly disagree) to 6

(strongly agree). Therefore, the variables used in the survey are continuous, and corresponding machine learning algorithms must be chosen for analysis. The algorithms that support continuous variables are called regression algorithms. The author of this study has chosen to use the following machine learning regression algorithms: Random Forest, XGBoost, and Linear Regression.

Random Forest and XGBoost are optimization frameworks for decision tree algorithms. The decision-making process is frequently represented as decision trees, which have a branching, tree-like form. The method involves branching decisions that lead to outcomes, creating a structure or visualization that resembles a tree. The key advantages of employing a decision tree in machine learning are its simplicity and the ease with which the decision-making process can be seen and understood. Pruning of the tree structure is frequently required because decision trees in machine learning can produce excessively complicated branches (Castillo, 2021).

All these algorithms use different approaches in determining the critical factors that affect a specific variable the most. Those approaches in machine learning are called feature importance methods. A more accurate picture can be achieved by applying different algorithms and methods. For instance, Random Forest and XGBoost algorithms are great for processing complex and multi-dimensional data and can detect the non-linear relationships between variables. On the other hand, the Linear Regression algorithm is a more straightforward and interpretable algorithm that can determine linear relationships between variables. When using several algorithms, complete comprehension can be achieved in determining which of the

factors significantly impacts overall job satisfaction (Breiman, 2001; Chen & Guestrin, 2016; Montgomery et al., 2012).

XGBoost. XGBoost (Extreme Gradient Boosting) is a machine learning optimization framework for decision trees used to predict a regression problem's outcome. It is an optimized gradient-boosting open-source package that is highly efficient, adaptable, and convenient. XGBoost works by building a sequence of decision trees iteratively, where each new tree tries to adjust the mistakes made by the preceding tree, resulting in a highly accurate model (Chen & Guestrin, 2016).

XGBoost has been broadly used in numerous fields, including finance, healthcare, and marketing, due to its high accuracy and interpretability. It has been shown to outperform other machine learning algorithms, such as Random Forest, Support Vector Machines (SVM), and Artificial Neural Networks (ANN) in several benchmark tests (Chen & Guestrin, 2016). One of the advantages of XGBoost is its capacity to deal with missing and skewed data, which is common in real-world datasets.

In the context of job satisfaction, XGBoost can be used to identify the factors that have the most significant effect on employee job satisfaction by evaluating the relationships between different predictors and job satisfaction. This can help companies develop efficient approaches to increase job satisfaction and retain critical top-performing employees.

Random Forest. The Random Forest algorithm is a supervised machine learning optimization framework for decision trees mainly applied for classification and regression problems (Biau & Scornet, 2016). The algorithm mixes several randomized decision trees and averages their predictions (Y. Chen et al., 2022). The internal nodes of a decision tree are tests of attributes, the branches are the results of the tests, and the leaf nodes are decisions taken once all attributes have been analyzed. (Zhou, 2021). The Random Forest algorithm approach performs well when the number of predictors surpasses the number of observations.

Furthermore, it is a flexible algorithm that can solve large and complex problems and return a measure of the predictors' significance (Biau & Scornet, 2016). Chen et al. (2022) describe that for model setting, the Random Forest algorithm picks a random sample of the predictors and observations at every decision tree node as the training dataset. Nevertheless, due to the arbitrary selection of the training dataset, the model's results might differ; therefore, a good practice is to apply a large set of decision trees to assure the accuracy of the outcome (Y. Chen et al., 2022). Thus, the Random Forest algorithm can also identify the predictors that affect employee job satisfaction the most.

Linear Regression. According to Maulud & Abdulazeez (2020), the Linear Regression algorithm predicts future events based on the linear relationship between independent and dependent variables. The algorithm is widely used for forecasting, predicting, and identifying the correlation between dependent and independent variables. In the research by Choi and Choi in 2022, the Linear Regression algorithm was successfully used to predict job satisfaction. The algorithm predicts the value of a dependent variable Y (outcome) based on an independent variable X (input) (Maulud & Abdulazeez, 2020). Linear Regression shows the linear relationship, demonstrating how the dependent variable changes as the independent variable changes. Therefore, the Linear Regression algorithm is also suitable for identifying the predictors' effect on job satisfaction.

Methods to Determine Importance of Predictors

The author used three methods for determining the essential variables to get more comprehensive results on the research question. The first method is a default feature importance method provided by Sci-kit learn library. The second method that was used is a permutation feature importance method provided by the Sci-kit Learn library. The permutation method involves shuffling the data many times and calculating the decrease in a model score. The third method used is a drop-column feature importance, which deletes a column from the dataset, retrains the model, and observes how it affects accuracy.

Default Feature Importance Method. The default feature importance method is based on the impurity reduction that a feature provides across all the trees in the model. The importance score is computed by averaging the impurity decrease over all the trees for each feature. A higher score means higher feature importance (Rajbahadur et al., 2022). However,

impurity-based feature importances might be ambiguous when using variables that have many unique values (*Permutation Importance vs Random Forest Feature Importance (MDI)*, n.d.).

Permutation Feature Importance Method. The permutation feature importance method randomly rearranges the values of each variable and measures the effect on the model's performance. If a feature is essential, shuffling its values should significantly decrease the model's prediction performance. This method is computationally expensive due to the sheer number of trials it generates but provides accurate feature importance scores (Altmann et al., 2010).

Drop-column Feature Importance Method. Drop-column feature importance method involves retraining the model after removing each feature and measuring the effect on the model's performance. If a feature is important, removing it should significantly decrease the model's performance. This method is also computationally expensive but provides accurate feature importance scores. However, in the scope of this research, the algorithms are processing small amounts of data; therefore, computational expenses are insignificant (Terence Parr et al., 2018). The author wrote this method using the logic and explanations available in the existing literature (Badeer, 2022; Shion Honda, 2020)

Performance Metrics. Data were split into training and testing sets during the data analysis to evaluate the quality and performance of the algorithms used in the research. The author of this study used two methods to estimate the performance metrics of the machine learning algorithm models. One method used in the research is the coefficient of determination (R^2), which calculates the amount of the target variable's variation that the model can account for. The closer the score to 1, the better. Another method the author uses is the Mean Squared

Error (MSE), which calculates the square root of the mean difference between the expected and actual values (Breiman, 2001). The closer the score to 0, the better.

The author coded all the technical parts in Python using the Google Colab product from Google Research. The author used the following libraries to complete the technical part: pandas, numpy, sklearn, matplotlib, and xgboost.

Identification of Job Satisfaction Predictors by Using Machine Learning Algorithms

The main goal of a Business is to be profitable, be able to compete, and thus be successful. The organization considers many variables, including the job satisfaction parameter, to reach the goal. Its effect on a company's profitability and performance varies, but it exists. Danish and Usman (2010) state that job satisfaction does not directly correlate with employee performance. However, rewards, competitive salaries, and a good working environment can improve job satisfaction and performance. With the help of machine learning algorithms, it becomes possible to identify which factors significantly impact job satisfaction, thus improving performance and, consequently, profits. In addition, employers could use the data to address problematic parts of their work structure, improving employee satisfaction and well-being, resulting in a win for both parties.

Another application of knowing such information is if the organization has no opportunities to improve the most critical factor. Every individual has unique factors that affect job satisfaction the most. Therefore, when recruiting newcomers, the employer can select those individuals whose satisfaction will not be affected by the most vulnerable part of the organization. This approach might be the best when the organization cannot improve critical

parts due to financial or other reasons. In such a scenario, employee satisfaction will be at the highest possible rate in such conditions, while the employer will hire a decent worker.

The study aims to determine the effect of predictors on job satisfaction. The following question will guide the research to find out what the most crucial factors are affecting job satisfaction and how could machine learning approach be used to identify these factors:

“Can the key factors that impact employee job satisfaction in the workplace be identified, and how can a machine learning approach be used to determine the impact?”

Method

Participants

The participants were selected using a convenience sampling method at company X, which specializes in audit and legal services. They include but are not limited to professionals such as financial and IT auditors, associates, managers, accountants, and advisors. The survey was sent to Latvian, Estonian, and Lithuanian departments to capture as many responses as possible. The study collected data from 128 participants, with 52% identifying as male and 48% identifying as female. The average age of the participants was 29 years old, ranging from 21 to 46 years old. All participants were informed about the purpose of the research before participating in the study, and their privacy and confidentiality were ensured throughout the research process.

Instrument

To obtain qualitative results that affect job satisfaction, the author used Paul Spector's Job Satisfaction Survey (JSS), which consisted of questions related to various aspects of job satisfaction, such as pay, benefits, job security, and work-life balance (Spector, 1985). The more factors covered in the research, the more specific answer will be in describing which of the predictors has the most significant effect on job satisfaction. According to Spector (1985), the Job Satisfaction Survey assesses nine dimensions linked to overall satisfaction with an average correlation coefficient of 0.70. The dimensions will be used as predictors in further study. Paul Spector's survey does not cover new modern trends such as work-from-home opportunities; nevertheless, all the predictors covered in the survey are relevant in modern organizations. Respondents were asked to answer the questions on a scale from 1 (strongly disagree) to 6

(strongly agree). An even number of answers forces the respondents to choose a definitive answer rather than allowing them to select a neutral or noncommittal response. This method can help to avoid ambiguity and provide more qualitative data for analysis. Additionally, an even number of answers prevents a respondent from simply selecting the middle option as a default, which can skew the results.

Procedure

The procedure for this study involved administering the Job Satisfaction Survey to a sample of 128 participants. The survey was created in Microsoft Teams in English and sent out via email. To collect the data, the survey was sent to 3 Baltic offices: Latvia, Estonia, and Lithuania. The contact emails were taken from the employer. The survey data was then collected and analyzed to identify patterns and relationships between the different survey questions and overall job satisfaction. The data obtained from the survey was used to train machine learning algorithms to determine which predictors have the most significant impact on job satisfaction. Specifically, the survey data was used to create a model that can predict job satisfaction based on the various predictors included in the survey and determine their significance. The model was trained using several machine learning algorithms, namely Random Forest, XGBoost, and Linear Regression, and the performance of each algorithm was estimated based on the Mean Squared Error (MSE) and R Squared (R^2) scores. The complete procedure consists of 5 major steps: design the questionnaire, gather the contact emails, send the survey, analyze data, and confirm results.

Data Preprocessing

A few computational manipulations are necessary to interpret the Job Satisfaction Survey results. The survey consists of questions written in both directions - positive and negative. The survey has nine sections, each with four questions, and the total score is based on all 36 questions. Scores on each section can range from 4 to 24, and the total score can range from 36 to 216. Each question is scored from 1 to 6. High scores on the survey mean that the person is satisfied with their job. If someone strongly agrees with a negative question, it is the same as if they strongly disagree with a positive question. To account for the negative wording in some questions, the totals for those questions must be reversed. The values for each original item score should be replaced with its corresponding reversed item score. After reversing negatively worded items, the scores must be linked with the corresponding predictor, and the values must be added up (Spector, 1985).

Results

Descriptive Statistics

Descriptive statistics were used to summarize the characteristics of the sample. The sample consisted of 128 respondents. Table 1 shows the distribution of the answers about demographics. It shows that gender distribution is almost equal, with 48% identifying as female and 52% identifying as male. Most respondents (62%) were in the age group 18 to 30, while 33% were in the age group 31 to 40, and 5% were in the age group 41 to 50. Most respondents (66%) had only a bachelor's degree, 29% had a master's degree, and 5% had an unfinished bachelor's degree. The table also shows that most respondents (78.9%) evaluated their overall satisfaction level from 4 to 5, meaning that they are more satisfied rather than not satisfied.

Table 1

Demographics, the total number of respondents = 128

	N	%		N	%
Gender			Satisfaction level		
Female	62	48.44%	1	1	0.78%
Male	66	51.56%	2	5	3.91%
			3	14	10.94%
Age			4	53	41.40%
18-30	79	61.72%	5	48	37.50%
31-40	42	32.81%	6	7	5.47%
41-50	7	5.47%			
Education level					
Bachelor	85	66.41%			
Master	37	28.91%			
Unfinished bachelor	6	4.68%			

Note. Data is compiled from the author's research.

The figures from Table 2 provide important information about the distribution of the data in the sample for age, overall job satisfaction level, and already computed values for each predictor using Python. The sample's standard deviation shows the scattered data points around the mean. A low standard deviation indicates that the data points are close to the mean, whereas a significant standard deviation indicates that the data points are widely dispersed. By looking at Table 2, it can be learned that the standard deviation for satisfaction level is relatively low at 0.93. This can be explained by the fact that most respondents (78.9%) evaluated their overall satisfaction level from 3 to 4. Also, this table shows that Supervision and Co-workers predictors have the highest mean values (20.58 and 19.52, respectively); however, to get a more accurate answer about what affects job satisfaction the most, an in-depth analysis of the sample is necessary.

Table 2

Means and standard deviations for the age, overall job satisfaction, and predictors, the total number of respondents = 128

	Mean	Standard deviation
Age	28.84	6.88
Satisfaction level	4.27	0.93
Pay	15.17	2.73
Promotion	15.32	2.34
Supervision	20.58	2.66
Fringe benefits	16.95	3.24
Contingent rewards	18.66	3.01
Operating procedures	15.12	3.17
Co-workers	19.52	2.77
Nature of work	16.77	3.51
Communication	18.23	2.75

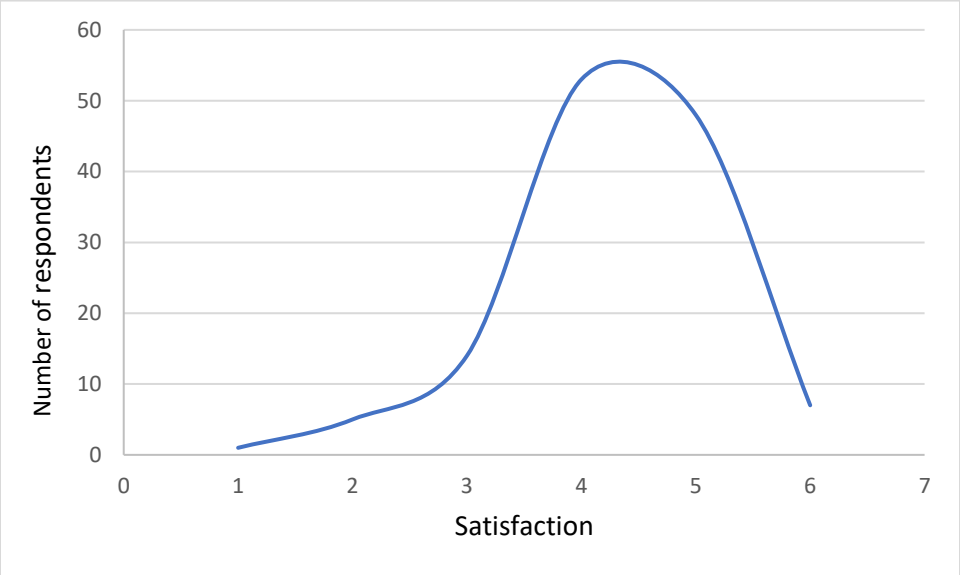
Note. Data is compiled from the author's research.

Inferential Statistics

Figure 1 shows that the dataset is almost normally distributed with a slight skewness to the right, meaning that most respondents are satisfied rather than dissatisfied. Therefore, there is no need to use any data transformation techniques, as all the beforementioned algorithms can handle such data distribution.

Figure 1

Distribution among the respondents regarding their overall job satisfaction level



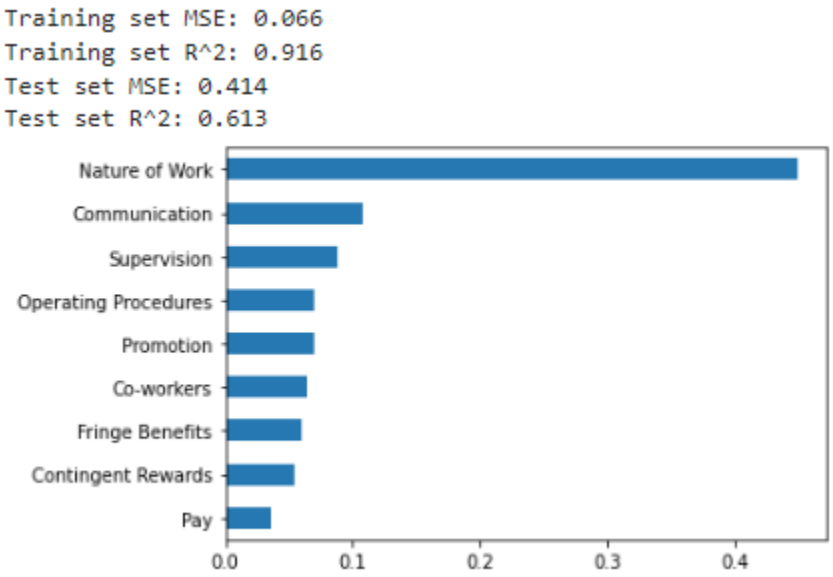
Random Forest

Random Forest algorithm was used to determine which predictors most affect overall job satisfaction. Random Forest offers classification and regression algorithms depending on the dataset. Since the dataset obtained from the survey contains continuous variables, the regression algorithm was used. The first method used to determine the impact of the predictors on overall job satisfaction is a default functionality in the Scikit-learn library in Python. Figure 2

shows the results obtained from using the default Scikit-learn library for feature importance determination in Python. The figure shows that the "Nature of Work" predictor significantly affects job satisfaction. The R^2 score on the training data was good (92%), while the MSE score of 0.414 on the test data and the R^2 score on the test data of 0.613 indicate that the model is not accurately predicting the outcome variable. An R^2 value of 0.613 indicates that the model only efficiently explains 61% of the variance in the outcome variable. The MSE score of 0.414 indicates that the model's predictions have relatively high error rates.

Figure 2

Default Scikit-learn Random Forest feature importance method



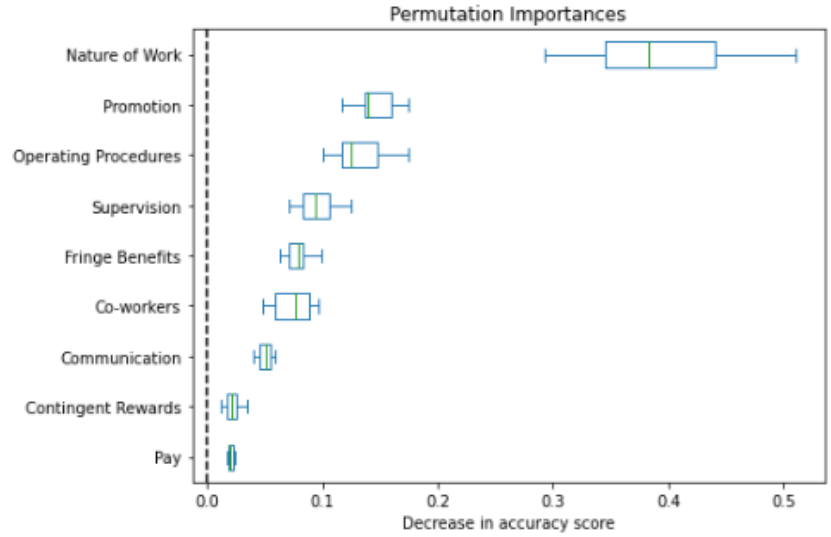
The permutation method is the second method for evaluating the feature importance using the Random Forest algorithm. Figure 3 shows the Sci-kit learn permutation feature importance method results. The figure shows that the “Nature of Work” variable is still the most crucial predictor with the most significant impact on overall job satisfaction. The value of 0.391 means that by randomly shuffling the values of the “Nature of Work” predictor, the

model's accuracy decreases by 39.1%. At the same time, +- 0.063 is the standard deviation of the mean feature importance score.

Figure 3

Random Forest Permutation Scikit-learn feature importance method

```
Nature of Work       0.391 +/- 0.063
Promotion            0.145 +/- 0.017
Operating Procedures 0.130 +/- 0.023
Supervision          0.095 +/- 0.016
Fringe Benefits      0.078 +/- 0.011
Co-workers           0.074 +/- 0.016
Communication        0.050 +/- 0.006
Contingent Rewards   0.022 +/- 0.006
Pay                  0.021 +/- 0.002
```



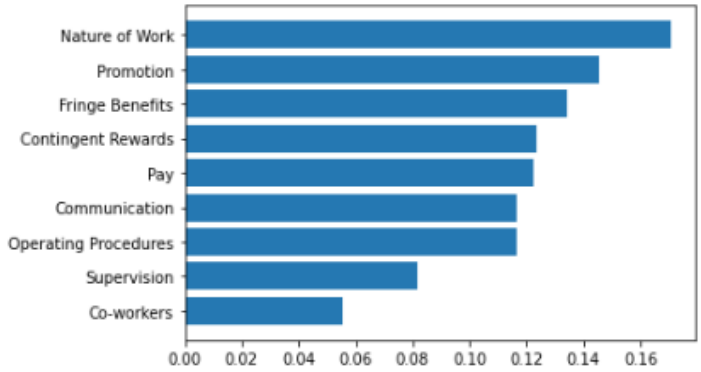
Drop-column feature importance is the third method used in the research for determining the most critical predictors affecting job satisfaction. In Figure 4, the importance scores can be observed. The figure shows that "Nature of Work" has the highest scores, followed by "Promotion" and "Fringe Benefits." The model suggests that these are 3 of the essential predictors affecting job satisfaction the most. However, from the figure, it can be learned that variables do not have significant importance score differences. Seven variables' scores range from 0.11 to 0.17, which means that variation in the outcome variable may not be

strongly influenced by a single factor but rather by a combination of variables. Interaction between predictors in the model must be considered, not only the individual predictor's score.

Figure 4

Random Forest drop-column feature importance method

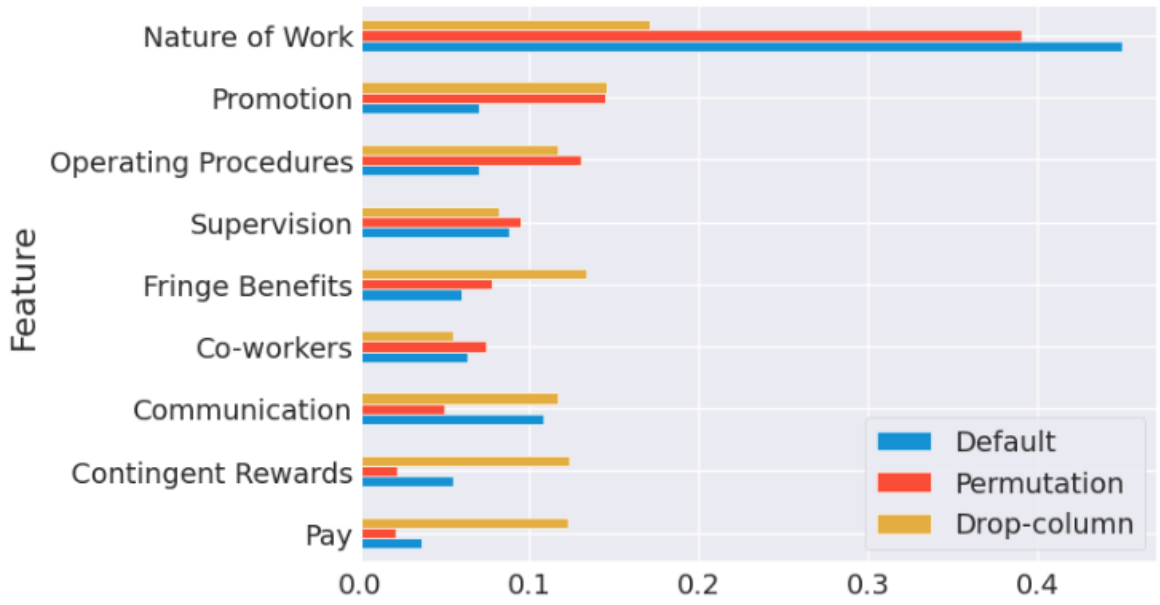
Co-workers	0.055185
Supervision	0.081649
Operating Procedures	0.116321
Communication	0.116606
Pay	0.122639
Contingent Rewards	0.123780
Fringe Benefits	0.134042
Promotion	0.145442
Nature of Work	0.170916



By combining all three Random Forest feature importance methods on one figure, a more comprehensive understanding of the factors that affect job satisfaction can be observed. Figure 5 shows the feature importance ranking of each variable from highest to lowest. Blue bars are the default feature importance method's scores, red bars are the permutation feature importance method's scores, and orange bars are the drop-column feature importance method's scores. Overall, the figure shows that the "Nature of Work" and "Promotion" variables are consistently the most important factors that have the most significant effect on job satisfaction.

Figure 5

Random Forest default, permutation and drop-column feature importance methods' ranking



XGBoost

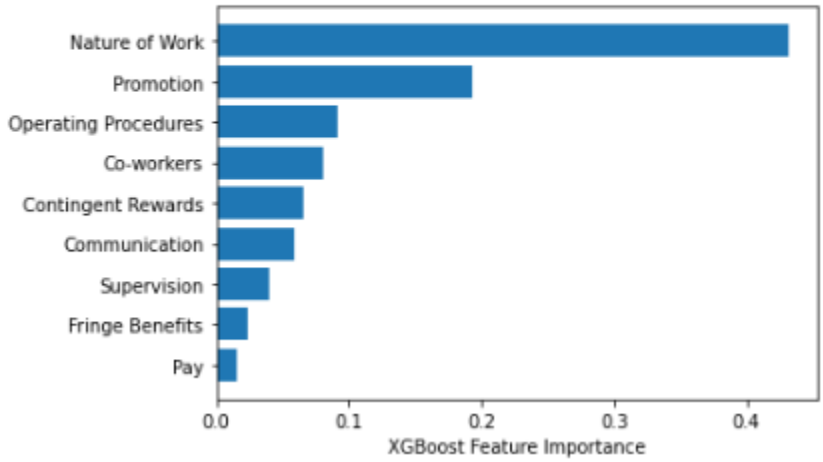
XGBoost provides a feature importance score that can be used to identify which variables have the greatest effect on a specific variable. The first method's results can be observed in Figure 6. A default XGBoost library with a feature importance method was used to obtain such results. From the bar, it can be observed that the "Nature of Work" is the most crucial factor with importance of over 40%. However, the XGBoost tends to overfit small datasets, so the algorithm, unfortunately, cannot perform accurately with new unseen data. The algorithm performed well on train data with an R^2 score of 96.8% and an MSE score of 2.9%, while with the test data, the R^2 score dropped to 20.1%, and the MSE score dropped to 50.2%. The R^2 score of 20.1% suggests that the model explains only 20.1% of the variability in

the outcome variable. Similarly, the MSE score of 50.2% indicates that the model makes relatively large prediction errors.

Figure 6

Default XGBoost feature importance method

```
Training set MSE: 0.029
Training set R^2: 0.968
Test set MSE: 0.502
Test set R^2: 0.201
```

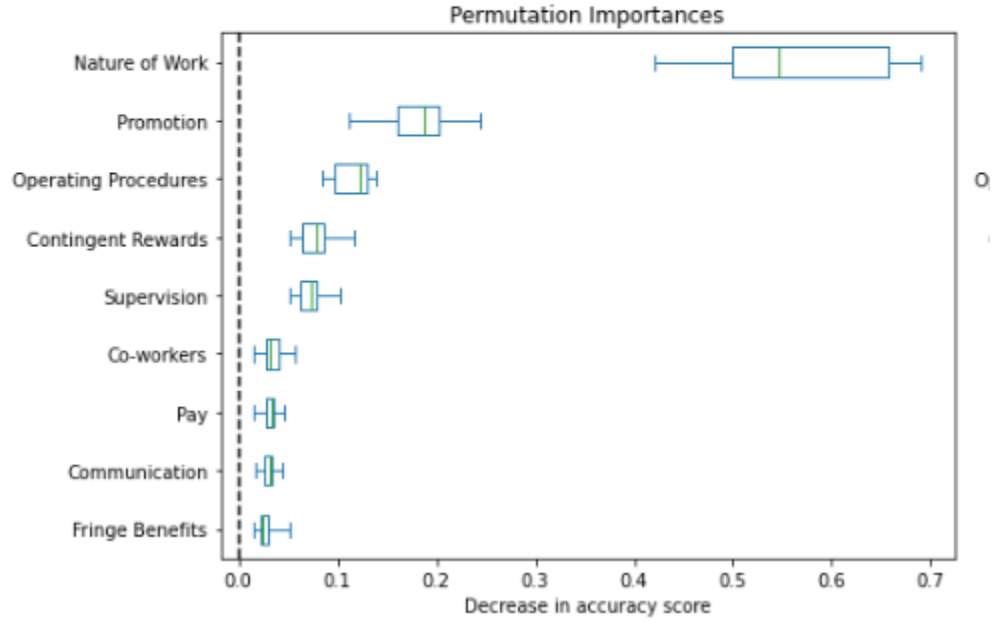


The permutation method is the second method for evaluating the feature importance using the XGBoost algorithm. Figure 7 shows the results obtained from using Sci-kit learn permutation feature importance method. The figure shows that the “Nature of Work” is the most critical predictor with the most significant impact on overall job satisfaction. The value of 0.564 means that by randomly shuffling the values of the “Nature of Work” predictor, the model’s accuracy decreased by 56.4%. The algorithm states that the least significant predictor is “Fringe Benefits.” According to the model, by randomly shuffling the values of “Pay,” the model’s accuracy decreased by 2.7%.

Figure 7

XGBoost Sci-kit learn permutation feature importance method

```
Nature of Work       0.564 +/- 0.085
Promotion            0.184 +/- 0.033
Operating Procedures 0.114 +/- 0.018
Contingent Rewards   0.078 +/- 0.018
Supervision          0.073 +/- 0.013
Co-workers           0.035 +/- 0.011
Pay                  0.032 +/- 0.007
Communication        0.030 +/- 0.007
Fringe Benefits      0.027 +/- 0.008
```

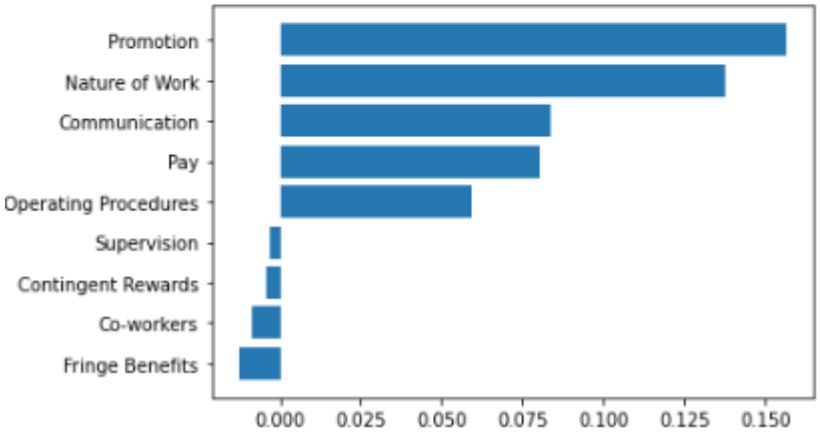


The third method used to determine the importance of predictors using the XGBoost algorithm is drop-column feature importance. Figure 8 shows the results obtained by using the third method. The figure shows that “Promotion” and “Nature of Work” predictors significantly affect job satisfaction. By dropping the “Promotion” predictor, the model’s accuracy decreased by 15.6%. The method has identified that the “Supervision,” “Contingent Rewards,” “Co-workers,” and “Fringe Benefits” variables have a negative score. The model’s accuracy improved by dropping those variables, which means they are irrelevant. As a result, the algorithm is performing better without these variables.

Figure 8

XGBoost drop-column feature importance method

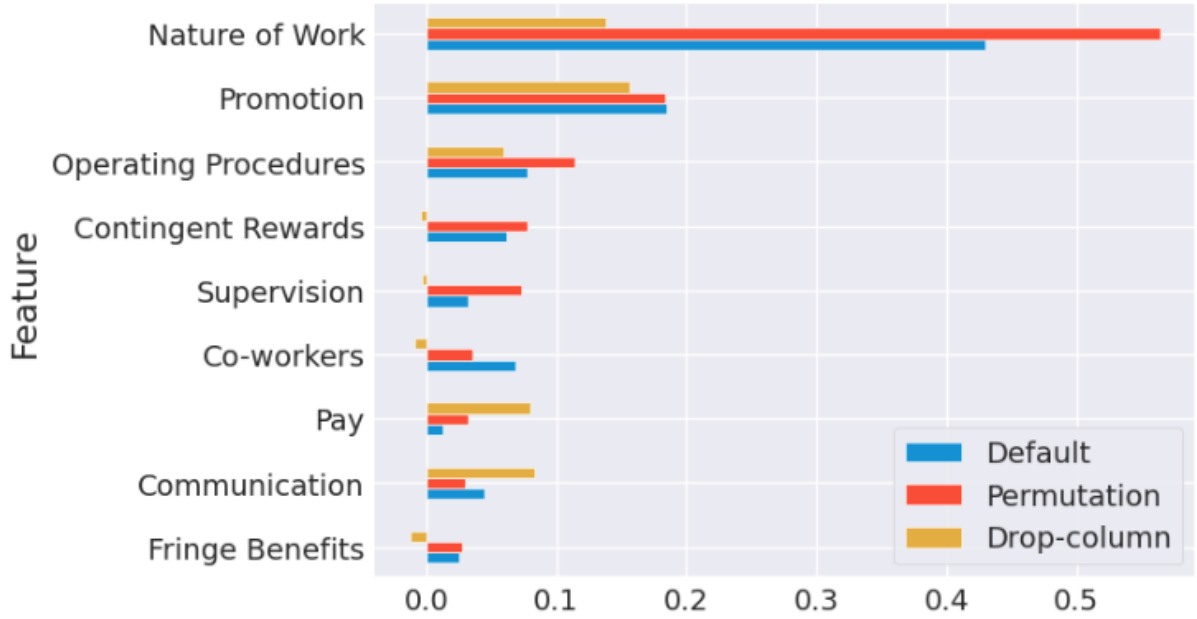
Fringe Benefits	-0.012669
Co-workers	-0.008696
Contingent Rewards	-0.004261
Supervision	-0.003100
Operating Procedures	0.059370
Pay	0.080557
Communication	0.083650
Nature of Work	0.137922
Promotion	0.156667



Combining all three XGBoost feature importance methods on one figure allows a more comprehensive understanding of the factors that affect job satisfaction. Figure 9 shows the feature importance ranking of each variable from highest to lowest. Blue bars are the default feature importance method's scores, red bars are the permutation feature importance method's scores, and orange bars are the drop-column feature importance method's scores. Overall, the figure shows that the "Nature of Work" and "Promotion" variables are consistently the most important factors that have the most significant effect on job satisfaction.

Figure 9

XGBoost default, permutation and drop-column feature importance methods' ranking



Linear Regression

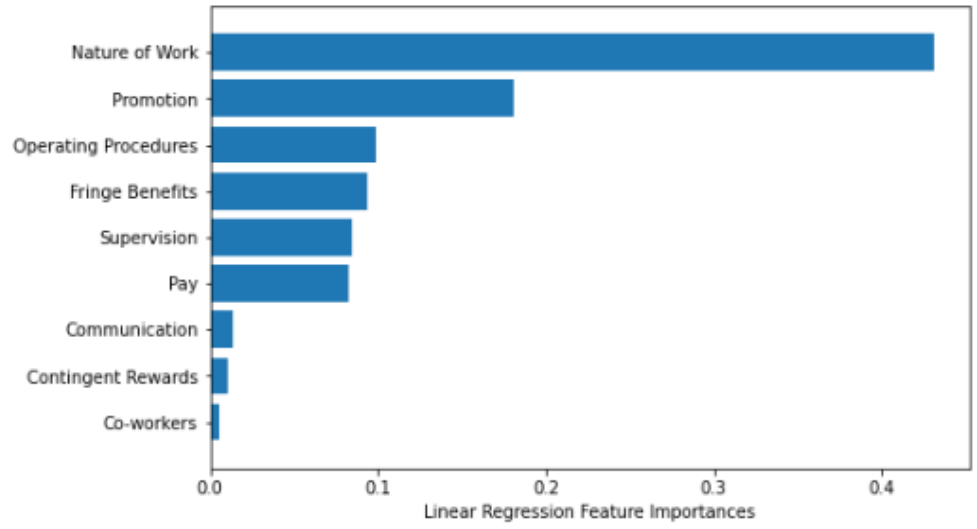
Linear Regression is a third machine learning algorithm that was used in the study to determine the importance of the variables on job satisfaction. The first method's results can be observed in Figure 10. A default Sci-kit learn library with a feature importance method was used to obtain such results. From the bar, we can observe that the "Nature of Work" is the most critical factor, with an importance of 0.431. However, the Linear Regression algorithm showed low scores on training and test data sets – the R^2 score on training and test sets are 0.588 and 0.394, respectively, and the MSE scores on training and test data sets are 0.377 and 0.381, respectively. The scores indicate that the model is making relatively large prediction errors. This can be explained by the small amount of data gathered for training.

Figure 10

Default Sci-kit learn Linear Regression feature importance method

```
Training R^2 score: 0.588
Training MSE score: 0.377
Test R^2 score: 0.394
Test MSE score: 0.381

Nature of Work: 0.431
Promotion: 0.181
Operating Procedures: 0.098
Fringe Benefits: 0.094
Supervision: 0.085
Pay: 0.082
Communication: 0.013
Contingent Rewards: 0.011
Co-workers: 0.005
```



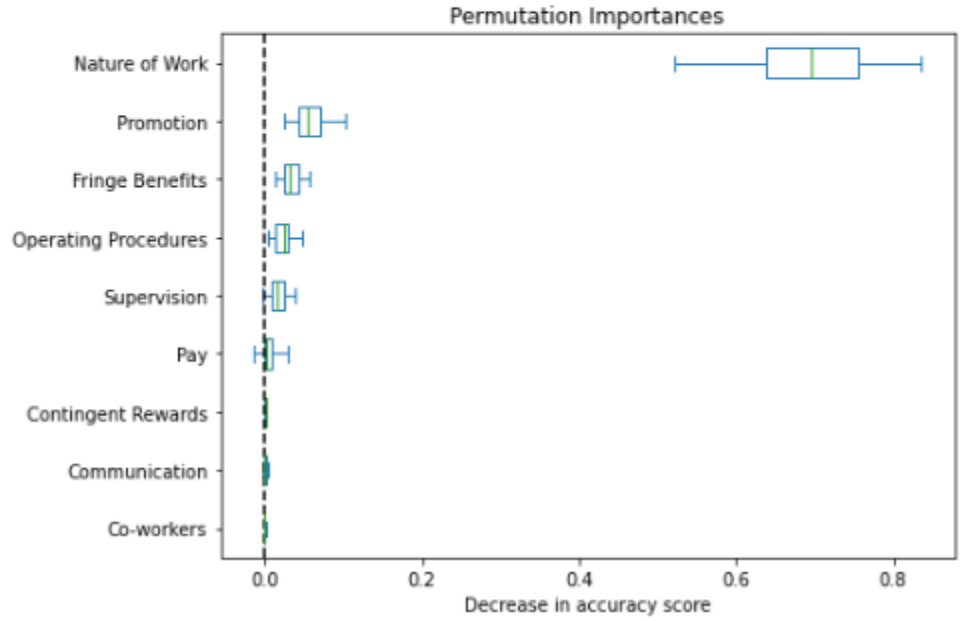
Using a Linear Regression algorithm, the permutation method is the second technique for evaluating the feature importance. Figure 11 shows the results obtained from using Sci-kit learn permutation feature importance method. The figure reveals that the “Nature of Work” is the most significant predictor impacting overall job satisfaction. The value of 0.694 means that by randomly shuffling the values of the “Nature of Work” predictor, the model’s accuracy

decreases by 69.4%. The algorithm states that the least significant predictor is “Co-workers,” with almost no effect.

Figure 11

Sci-kit learn Linear Regression permutation feature importance method

```
Nature of Work       0.694 +/- 0.088
Promotion            0.057 +/- 0.021
Fringe Benefits      0.034 +/- 0.013
Operating Procedures 0.023 +/- 0.011
Supervision          0.017 +/- 0.010
Pay                  0.005 +/- 0.010
Contingent Rewards   0.001 +/- 0.001
Communication        0.000 +/- 0.002
Co-workers           0.000 +/- 0.001
```



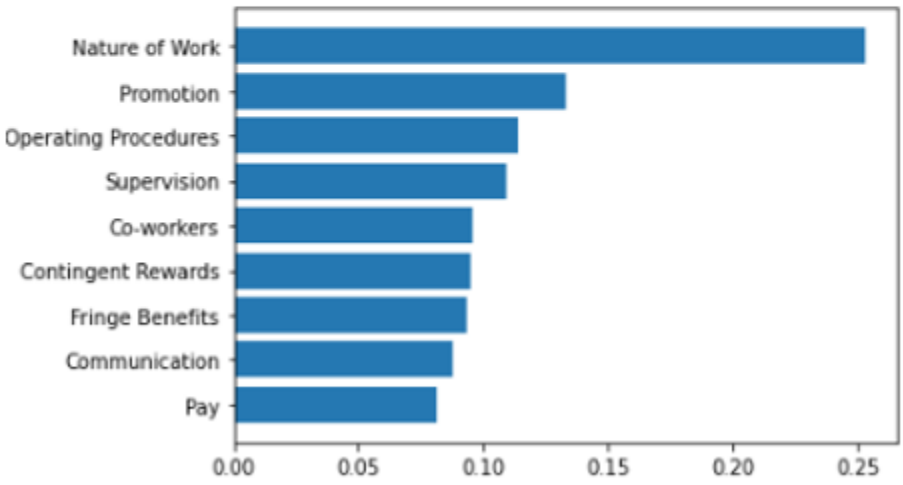
The third method used to determine the importance of the factors affecting job satisfaction using the Linear Regression algorithm is the drop-column feature importance method. From Figure 12, it can be learned that the “Nature of Work” has the most significant impact on job satisfaction. By dropping this predictor, the model’s accuracy decreased by 25.3%. The next most important variable is “Promotion,” with an importance score of 0.133.

The gap between the first and second positions is extensive, while other factors' scores do not significantly differ. This can be explained by the fact that the "Nature of Work" has a more substantial influence than other factors, which have approximately the same impact on job satisfaction.

Figure 12

Linear Regression drop-column feature importance method

Pay	0.081303
Communication	0.087759
Fringe Benefits	0.093702
Contingent Rewards	0.095263
Co-workers	0.095637
Supervision	0.108913
Operating Procedures	0.114270
Promotion	0.133418
Nature of Work	0.253479

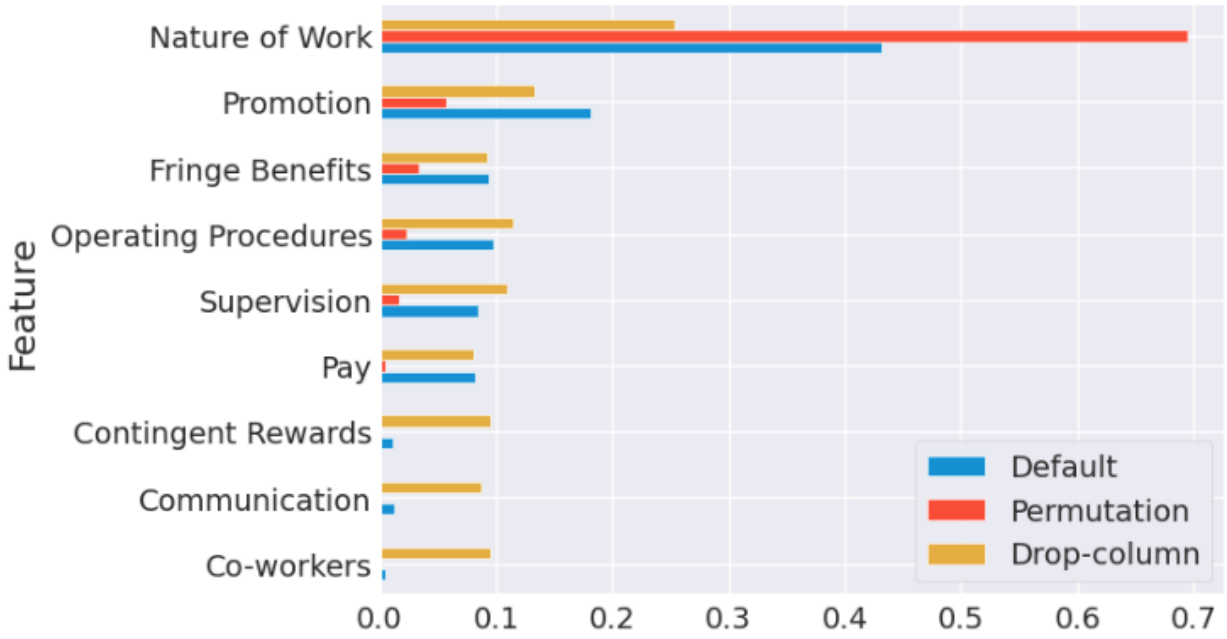


By combining all three Linear Regression feature importance methods on one figure, a more comprehensive understanding of the factors that affect job satisfaction can be observed. Figure 13 shows the feature importance ranking of each variable from highest to lowest. Blue bars are the default feature importance method's scores, red bars are the permutation feature

importance method's scores, and orange bars are the drop-column feature importance method's scores. Overall, the figure shows that the "Nature of Work" and "Promotion" variables are consistently the most important factors that have the most significant effect on job satisfaction.

Figure 13

Linear Regression default, permutation and drop-column feature importance methods' ranking



Evaluation

Data were split into training and test sets during the data analysis to evaluate the quality and performance of the algorithms used in the research. Further, the MSE and R^2 values were calculated for each algorithm on the testing data sets. The results are summarized in Table 3 below. Table 3 shows that the Random Forest algorithm performed the best among the other algorithms, with the lowest MSE and highest R^2 values. Meanwhile, the XGBoost

algorithm had the highest MSE and lowest R^2 values, indicating that it was the least accurate algorithm for predicting job satisfaction.

Table 3

Performance metrics for each algorithm

	Mean Squared Error (MSE)	R-squared (R^2)
Random Forest	0.41	0.61
XGBoost	0.50	0.20
Linear Regression	0.38	0.39

Note. Data is compiled from the author's research.

Even though the MSE scores were relatively high and R^2 scores were relatively low, the research question is answered. Indeed, it is possible to determine the impact of each predictor on overall job satisfaction using machine learning algorithms.

Discussion

The research provides valuable insights into the determinants of job satisfaction, particularly the importance of features, using machine learning algorithms. To achieve this, three algorithms - Random Forest, XGBoost, and Linear Regression, and three different methods were applied to compute the predictive importance of each variable: the default feature importance method, permutation importance method, and drop-column feature importance method.

The research results revealed that across all three algorithms and three feature importance methods, "Nature of Work" and "Promotion" were consistently identified as the two most significant job satisfaction factors. The "Nature of Work" became the most critical variable overall. This result is consistent with prior research that emphasizes the role of job characteristics in job satisfaction, especially job nature and professional growth prospects (Hackman & Oldham, 1975; Judge et al., 2001).

It is worth noting that the default feature importance method with the Random Forest algorithm identified "Communication" as the second most important feature, while "Nature of Work" remains the most important feature. This discrepancy may be due to how the default feature importance method works, as it is based on decreasing the impurity of each feature in the decision tree. Consequently, it may not capture the full complexity of the relationship between traits and job satisfaction. This could explain why “Communication” was not identified as highly important in other models or methods. Another discrepancy was identified using the XGBoost algorithm drop-column feature importance method. In this case, the “Promotion” variable was determined to have the most significant impact on job satisfaction, followed by

“Nature of Work.” The disparity might be because XGBoost is a gradient boost optimization framework for decision trees that can identify a complex nonlinear relationship between variables, while Random Forest and Linear Regression have limitations. There is also a possibility that the dataset used in the research has biases, and the algorithms used in the dataset interpret those biases differently. Therefore, several methods are necessary to cross-validate the results.

Hackman and Oldham's (1975) job characteristics model identified five key characteristics—skill variety, task identity, task significance, autonomy, and feedback—that influence job satisfaction and motivation the most. Similarly, (Judge et al., 2001) found that job characteristics such as skill variety, task identity, and task importance were significant predictors of job satisfaction. These relationships were related to psychological meaning and responsibility.

Previous research has also confirmed the importance of promotion prospects in job satisfaction. For example, (Judge et al., 2010) found that promotion opportunities positively correlate to job satisfaction and that perceptions of distributive justice partially mediate this relationship. In addition, (Kalleberg et al., 2000) found that promotion was positively correlated to job satisfaction and that this relationship was stronger for women than men.

Overall, the findings are consistent with previous research and emphasize the importance of job characteristics and professional development opportunities in determining employee job satisfaction. Using machine learning algorithms to determine feature importance can provide valuable insights for organizations trying to improve employee satisfaction. The

results suggest that multiple algorithms and methods should be used to ensure the importance of different variables.

Practical and Theoretical Implications

This research provides valuable insights for organizations aiming to improve employee job satisfaction. The study results show that specific job characteristics, such as “Nature of Work” and “Promotion”, significantly impact job satisfaction. Therefore, organizations must explore methods to improve these aspects to promote employee satisfaction.

Organizations can offer employees more freedom, opportunities for professional development, and meaningful and stimulating tasks to improve the nature of the work aspect. This can be achieved by restructuring job roles, providing training and development opportunities, and ensuring employees are engaged in work matching their skills and interests. In addition, organizations can improve opportunities for advancement by creating transparent career paths, mentoring and coaching programs, and clear promotion criteria. These initiatives can help employees feel valued and supported, increasing their commitment to the organization.

Research findings also highlight the importance of using different algorithms and methods to identify key factors. This will ensure that the identified predictors are reliable, allowing organizations to develop targeted interventions to improve employee satisfaction. Additionally, machine learning algorithms can provide a complete conception of the complex relationships between job characteristics and satisfaction, leading to more effective strategies for promoting job satisfaction.

In conclusion, the research underlines the importance of job characteristics and growth opportunities in determining employee job satisfaction and highlights the need for organizations to invest in these areas to attract and retain top talents. Such investments can create a more positive work environment and improve job satisfaction.

Research Limitations and Future Research Possibilities

Although the research has provided valuable insights, it is essential to note that some limitations should be considered. First, the sample size could have been bigger, so it may not accurately reflect the wider population. Also, it does not apply to different industries since the reviewed organization is specific regarding employee satisfaction. Therefore, it is vital to be cautious when concluding, generalizing results, and using them under different circumstances. Fundamental point limitations are data, selected predictors, and algorithms.

Reliance on self-reported data. Results may be affected by biases and may not accurately reflect the experiences and perceptions of the participants. To overcome this limitation, future research could incorporate a mixed methods approach that combines self-reported data with objective measures and observations.

Focus on a specific set of predictors. The research did not consider other factors, such as leadership and hybrid working mode, that may also affect job satisfaction. Future research could extend the parameters to consider a broader range of predictors and provide a more comprehensive understanding of the factors influencing job satisfaction.

Machine learning algorithms. The research used only three machine learning algorithms and three feature importance methods and did not try other algorithms or methods that might

produce different results. Future studies may use alternative algorithms and methods to ensure precise and generalizable findings.

Despite these limitations, the study has provided valuable insight into factors influencing job satisfaction and highlights the importance of using machine learning algorithms to determine feature importance. Future research could build up on these findings, addressing the abovementioned limitations and offering a more nuanced understanding of the complex relationships between predictors and job satisfaction.

References

Aguinis, H., Joo, H., & Gottfredson, R. K. (2013). What monetary rewards can and cannot do: How to show employees the money. *Business Horizons*, *56*(2), 241–249. https://doi.org/10.1016/j.bushor.2012.11.007

Altmann, A., Toloşi, L., Sander, O., & Lengauer, T. (2010). Permutation importance: A corrected feature importance measure. *Bioinformatics*, *26*(10), 1340–1347. https://doi.org/10.1093/bioinformatics/btq134

Badeer, P. (2022, April 5). Feature Importance May Be Lying To You. *Medium*. https://towardsdatascience.com/feature-importance-may-be-lying-to-you-3247cafa7ee7

Biau, G., & Scornet, E. (2016). A Random Forest guided tour. *TEST*, *25*(2), 197–227. https://doi.org/10.1007/s11749-016-0481-7

Castillo, D. (2021, November 13). Decision Trees in Machine Learning Explained. *Seldon*. https://www.seldon.io/decision-trees-in-machine-learning

Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. https://doi.org/10.1145/2939672.2939785

Chen, Y., Zhang, X., Lu, L., Wang, Y., Liu, J., Qin, L., Ye, L., Zhu, J., Shia, B.-C., & Chen, M.-C. (2022). Machine Learning Methods to Identify Predictors of Psychological Distress. *Processes*, *10*(5), 1030. https://doi.org/10.3390/pr10051030

Choi, Y., & Choi, J. (2022). *Job Satisfaction Prediction and Machine Learning Technique*. https://doi.org/10.21203/rs.3.rs-1683972/v1

Danish, R. Q., & Usman, A. (2010). Impact of Reward and Recognition on Job Satisfaction and Motivation: An Empirical study from Pakistan. *International Journal of Business and Management*, *5*(2). https://www.academia.edu/32872929/Impact_of_Reward_and_Recognition_on_Job_Satisfaction_and_Motivation_An_Empirical_study_from_Pakistan

De Togni, G., Erikainen, S., Chan, S., & Cunningham-Burley, S. (2021). What makes AI 'intelligent' and 'caring'? Exploring affect and relationality across three sites of intelligence and care. *Social Science & Medicine*, *277*, 113874. https://doi.org/10.1016/j.socscimed.2021.113874

den Dulk, L., Peters, P., & Poutsma, E. (2012). Variations in adoption of workplace work–family arrangements in Europe: The influence of welfare-state regime and organizational characteristics. *The International Journal of Human Resource Management*, *23*(13), 2785–2808. https://doi.org/10.1080/09585192.2012.676925

Galanaki, E. (2013). Gender and the Importance of Fringe Benefits: Exploring Their Link with Organizational Commitment and Job Satisfaction. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2325426

Hackman, J. R., & Oldham, G. R. (1975). Development of the Job Diagnostic Survey. *Journal of Applied Psychology*, *60*(2), 159–170. https://doi.org/10.1037/h0076546

Harper, E., Castrucci, B. C., Bharthapudi, K., & Sellers, K. (2015). Job Satisfaction: A Critical, Understudied Facet of Workforce Development in Public Health. *Journal of Public Health Management and Practice*, *21*(Supplement 6), S46–S55. https://doi.org/10.1097/PHH.0000000000000296

Huang, L. V., & Liu, P. L. (2017). Ties that work: Investigating the relationships among coworker connections, work-related Facebook utility, online social capital, and employee outcomes. *Computers in Human Behavior*, *72*, 512–524. https://doi.org/10.1016/j.chb.2017.02.054

IBM Cloud Education. (2022, July 6). *What is Machine Learning?* https://www.ibm.com/cloud/learn/machine-learning

Judge, T. A., Piccolo, R. F., Podsakoff, N. P., Shaw, J. C., & Rich, B. L. (2010). The relationship between pay and job satisfaction: A meta-analysis of the literature. *Journal of Vocational Behavior*, *77*(2), 157–167. https://doi.org/10.1016/j.jvb.2010.04.002

Judge, T. A., Thoresen, C. J., Bono, J. E., & Patton, G. K. (2001). The job satisfaction–job performance relationship: A qualitative and quantitative review. *Psychological Bulletin*, *127*(3), 376–407. https://doi.org/10.1037/0033-2909.127.3.376

Kalleberg, A. L., Reskin, B. F., & Hudson, K. (2000). Bad Jobs in America: Standard and Nonstandard Employment Relations and Job Quality in the United States. *American Sociological Review*, *65*(2), 256. https://doi.org/10.2307/2657440

Knoop, R. (1995). Relationships Among Job Involvement, Job Satisfaction, and Organizational Commitment for Nurses. *The Journal of Psychology*, *129*(6), 643–649. https://doi.org/10.1080/00223980.1995.9914935

Lerouge, C., Klaus, T., & Blanton, J. (2014). System Developers' Nature of Work Characteristics and Their Relationship with Organizational Commitment and Job Satisfaction. *Journal of Internet Technology Management*. http://jitm.ubalt.edu/XXV-1/article1.pdf

Mahesh, B. (2019). *Machine Learning Algorithms -A Review*. https://doi.org/10.21275/ART20203995

Maulud, D., & Abdulazeez, A. M. (2020). A Review on Linear Regression Comprehensive in Machine Learning. *Journal of Applied Science and Technology Trends*, *1*(4), 140–147. https://doi.org/10.38094/jastt1457

McNeese-Smith, D. (1996). Increasing Employee Productivity, Job Satisfaction, and Organizational Commitment. *Journal of Healthcare Management*, *41*(2), 160–175.

Pergamit, M. R., & Veum, J. R. (1999). What is a Promotion? *ILR Review*, *52*(4), 581–601. https://doi.org/10.1177/001979399905200405

Permutation Importance vs Random Forest Feature Importance (MDI). (n.d.). Scikit-Learn. Retrieved 10 April 2023, from https://scikit-learn/stable/auto_examples/inspection/plot_permutation_importance.html

Peterson, D. K., Puia, G. M., & Suess, F. R. (2003). 'Yo Tengo La Camiseta (I Have the Shirt On)': An Exploration Of Job Satisfaction and Commitment Among Workers In Mexico. *Journal of Leadership & Organizational Studies*, *10*(2), 73–88. https://doi.org/10.1177/107179190301000208

Qasim, S., Azam, C. F., & Nadeen, A. S. (2012). Exploring factors affecting employees' job satisfaction at work. *Journal of Management and Social Sciences*, *8*, 31–39. https://doi.org/10.46745/ilma.jbs.2012.08.01.04

Rajbahadur, G. K., Wang, S., Oliva, G. A., Kamei, Y., & Hassan, A. E. (2022). The Impact of Feature Importance Methods on the Interpretation of Defect Classifiers. *IEEE*

Transactions on Software Engineering, *48*(7), 2245–2261. https://doi.org/10.1109/TSE.2021.3056941

Raziq, A., & Maulabakhsh, R. (2015). Impact of Working Environment on Job Satisfaction. *Procedia Economics and Finance*, *23*, 717–725. https://doi.org/10.1016/S2212-5671(15)00524-9

Rehman, K., Zia-Ur-Rehman, Saif, N., Khan, A. S., Nawaz, A., & Rehman, S. ur. (2013). Impacts of Job Satisfaction on Organizational Commitment: A Theoretical Model for Academicians in HEI of Developing Countries like Pakistan. *International Journal of Academic Research in Accounting, Finance and Management Sciences*, *3*(1), 80–89.

Rhee, S.-Y., Hur, W.-M., & Kim, M. (2017). The Relationship of Coworker Incivility to Job Performance and the Moderating Role of Self-Efficacy and Compassion at Work: The Job Demands-Resources (JD-R) Approach. *Journal of Business and Psychology*, *32*(6), 711–726. https://doi.org/10.1007/s10869-016-9469-2

Sageer, A., Dr. Rafat, S., & Ms. Agarwal, P. (2012). Identification of Variables Affecting Employee Satisfaction and Their Impact on the Organization. *IOSR Journal of Business and Management*, *5*(1). https://doi.org/10.9790/487x-0513239

Schlechter, A., Thompson, N. C., & Bussin, M. (2015). Attractiveness of non-financial rewards for prospective knowledge workers: An experimental investigation. *Employee Relations*, *37*(3), 274–295. https://doi.org/10.1108/ER-06-2014-0077

Shion Honda. (2020, September 5). Different Measures of Feature Importance Behave Differently. *Hippocampus's Garden*. https://hippocampus-garden.com/feature_importance/

Spector, P. E. (1985). Measurement of human service staff satisfaction: Development of the Job Satisfaction Survey. *American Journal of Community Psychology*, *13*(6), 693–713. https://doi.org/10.1007/BF00929796

Spielberger, C. D. (Ed.). (2004). *Encyclopedia of applied psychology* (1st ed). Elsevier / Academic Press.

Terence Parr, Kerem Turgutlu, Christopher Csiszar, & Jeremy Howard. (2018, March 26). *Beware Default Random Forest Importances*. Explained.AI. http://explained.ai/decision-tree-viz/index.html

Valentine, S. R., Mathis, R. L., Jackson, J. H., & Meglich, P. A. (2020). *Human resource management* (Sixteenth edition). Cengage.

Wright, J. D. (2015). *International encyclopedia of the social & behavioral sciences* (2nd edition). Elsevier.

Zhou, Z.-H. (2021). *Machine Learning*. Springer Singapore. https://doi.org/10.1007/978-981-15-1967-3

Appendix

Questionnaire

1. Age
2. Gender
    a. Man
    b. Woman
    c. Non-binary
    d. Prefer not to say
3. Education level
    a. Bachelor
    b. Master
    c. Ph.D.
    d. Other…

		Disagree very much	Disagree moderately	Disagree slightly	Agree slightly	Agree moderately	Agree very much
4	I am satisfied with my job.						
5	I feel I am being paid a fair amount for the work I do.	1	2	3	4	5	6
6	There is really too little chance for promotion on my job.	1	2	3	4	5	6
7	My supervisor is quite competent in doing his/her job.	1	2	3	4	5	6
8	I am not satisfied with the benefits I receive.	1	2	3	4	5	6
9	When I do a good job, I receive the recognition for it that I should receive.	1	2	3	4	5	6

10	Many of our rules and procedures make doing a good job difficult.	1	2	3	4	5	6
11	I like the people I work with.	1	2	3	4	5	6
12	I sometimes feel my job is meaningless.	1	2	3	4	5	6
13	Communications seem good within this organization.	1	2	3	4	5	6
14	Raises are too few and far between.	1	2	3	4	5	6
15	Those who do well on the job stand a fair chance of being promoted.	1	2	3	4	5	6
16	My supervisor is unfair to me.	1	2	3	4	5	6
17	The benefits we receive are as good as most other organizations offer.	1	2	3	4	5	6
18	I do not feel that the work I do is appreciated.	1	2	3	4	5	6
19	My efforts to do a good job are seldom blocked by red tape.	1	2	3	4	5	6
20	I find I have to work harder at my job because of the incompetence of people I work with.	1	2	3	4	5	6
21	I like doing the things I do at work.	1	2	3	4	5	6
22	The goals of this organization are not clear to me.	1	2	3	4	5	6
23	I feel unappreciated by the organization when I think about what they pay me.	1	2	3	4	5	6
24	People get ahead as fast here as they do in other places.	1	2	3	4	5	6
25	My supervisor shows too little interest in the feelings of subordinates.	1	2	3	4	5	6
26	The benefit package we have is equitable.	1	2	3	4	5	6
27	There are few rewards for those who work here.	1	2	3	4	5	6
28	I have too much to do at work.	1	2	3	4	5	6
29	I enjoy my coworkers.	1	2	3	4	5	6
30	I often feel that I do not know what is going on with the organization.	1	2	3	4	5	6
31	I feel a sense of pride in doing my job.	1	2	3	4	5	6
32	I feel satisfied with my chances for salary increases.	1	2	3	4	5	6
33	There are benefits we do not have that we should have.	1	2	3	4	5	6

34	I like my supervisor.	1	2	3	4	5	6
35	I have too much paperwork.	1	2	3	4	5	6
36	I don't feel my efforts are rewarded the way they should be.	1	2	3	4	5	6
37	I am satisfied with my chances for promotion.	1	2	3	4	5	6
38	There is too much bickering and fighting at work.	1	2	3	4	5	6
39	My job is enjoyable.	1	2	3	4	5	6
40	Work assignments are not fully explained.	1	2	3	4	5	6

Full code of the research can be found using the following link -

https://github.com/Boris2907/Employee_job_satisfaction_machine_learning

Guarantee

I understand that, in accordance with the Law of the Republic of Latvia regarding intellectual property rights I, Boriss Verbickis, am the owner of the copyright in this Bachelor Thesis and that the copyright rests with me unless I specifically assign it to another person. The data, definitions, citations that are taken from other sources are fully acknowledged in my work. Neither this work, nor any part of it, in one form or another has ever been handed in to some other commission and has never been published.

Signature

/Boriss Verbickis/

April 14, 2023