INTERNATIONAL JOURNAL OF SOCIAL SCIENCE, EDUCATIONAL, ECONOMICS, AGRICULTURE RESEARCH AND TECHNOLOGY (IJSET) | ISSN (E): 2827-766X
Website: https://ijset.org/ | Volumes 5 No. 4 (2026)

ENHANCING JOB FIT PREDICTION IN CORPORATIONS – A COMPARATIVE MACHINE LEARNING STUDY UTILIZING GRADIENT BOOSTING

Bondan Ari Wijaya[1*], Imam Yuadi[2]
Universitas Airlangga, Surabaya, Indonesia
[1]bondan.ari.wijaya-2025@pasca.unair.ac.id, [2]imam.yuadi@fisip.unair.ac.id
*) Corresponding Author

Received:03 /03/2026 | **Revised**: 15/03/2026 | **Accepted:** 05/04/2026 | **Published**: 11/04/2026

Abstract
The test results demonstrated how well the Gradient Boosting model could predict outcomes, with the model achieving the best performance metrics, such as an overall accuracy of 98% with 10-fold cross-validation. using group learning techniques to evaluate job fit. This remarkable performance was attained despite the organizational dataset's inherent class imbalance. Crucially, the model showed constant effectiveness in every aspect of job fit. The majority class, Perfect Match (98.8%), is divided into groups based on the difference between PeG and PoG. The minor groups, Overqualified (96.2%) and Underqualified (96.5%), are also divided into groups with strong accuracy and memory. "Jenjang - Main Grp "Text" and "PeG" are the two most important things that can tell you work fit," according to the feature importance analysis. These data give us a solid, objective basis for future talent management and placement decisions by clearly demonstrating that there are distinct, data-driven patterns in placing people in jobs at a company. Machine Learning, Job Fit, Human Resources, Gradient Boosting and Personnel Analytics.

Keywords: Machine Learning, Job Fit, Human Resources, Gradient Boosting, Personnel Analytics

I. INTRODUCTION

1.1 Background

The incorporation of objective organizational grading metrics more especially, the difference between Person Grade (PeG) and Position Grade (PoG) into a predictive machine learning framework is the main innovation of this study. In contrast to earlier research that mostly used self-reported survey data to evaluate person-job fit, this study makes use of real company HRIS data, offering a more reliable and objective model for extensive hiring. Additionally, by giving HR professionals interpretable decision boundaries, this study closes the gap between high-accuracy "black-box" algorithms and the requirement for organizational transparency. The phrase "job fit" refers to the alignment between an employee's abilities and the demands of their job, which is a major factor in how well a business performs. In government agencies, the ideal job match entails ensuring that public services are delivered promptly, which reduces employee dissatisfaction and lowers turnover costs. It is difficult for Indonesian corporations to place the right people in the right jobs because traditional methods of evaluating job suitability often rely on subjective assessments, manual evaluations, and bureaucratic procedures that can lead to biases and mismatches. Using organizational data to identify trends and forecast machine learning to identify the best places for staff shows promise.

1.2 Problem Statement

Businesses frequently encounter numerous issues with their current personnel placement practices, including:

- Increasingly complex skill requirements;
- Not making good use of workers' skills and abilities;
- Promotional decisions based on transfers and individual opinions;
- without the necessary resources to assess a candidate's suitability for a position;
- Decision-making procedures must be transparent and equitable.

Subjectivity in transfer and wage decisions:

The way that companies currently position their employees has significant issues because it is centered on personal opinion, which makes it difficult to use talent effectively and to see important decisions like promotions and moves. Data-based, impartial tools for job fit are needed because things are becoming more complex and assessments are required. However, most organizations lack predictive strategies to deal with differences proactively, such as the difference between Person Grade (PeG) and Position Grade (PoG). Therefore, it is crucial to develop systematic, objective models that can predict and correct these PeG-PoG differences. This will make managing people a clear, predictive science rather than a reactive, qualitative activity.

1.3 Research Objectives

The purpose of this study is to:

1. Develop machine learning models to classify jobs based on how PeG and PoG differ.
2. Examine and contrast the Random Forest results using techniques such as Decision Trees and Gradient Boosting.
3. Determine the most significant factors influencing companies and job fit in organizations.
4. Provide practical advice on data-driven HRM.

1.4 Research Contributions

The primary novelty of this research lies in the utilization of objective organizational grading metrics specifically the discrepancy between Person Grade (PeG) and Position Grade (PoG) as the core predictors for job fit. Unlike previous literature that predominantly relies on subjective self-reported survey data, this study utilizes actual transactional HRIS data, which minimizes inherent bias. Furthermore, this study offers a methodological contribution by bridging a high-accuracy 'black-box' algorithm (Gradient Boosting) with an interpretable model (Decision Tree). This approach not only ensures precise predictions but also provides transparent decision boundaries, enabling HR practitioners to make explainable and data-driven personnel placement decisions in large-scale corporate environments.

The effectiveness of ensemble machine learning in corporate HR analytics has been thoroughly validated in academic research, and this study has demonstrated the use of machine learning in corporate HR analytics, achieved remarkable predictive accuracy (98%) with actual organizational data, provided interpretable models for personnel decision-making, and created a framework for predictive job fit assessment in governmental entities. With a predicted 98% Accuracy in work fit evaluation utilizing authentic, intricate The efficacy of ensemble machine learning in corporate HR analytics has been thoroughly validated in academic research. It successfully integrates this predictive job fit assessment framework for immediate usage and growth into government and public sector organizations, and by identifying the most significant job fit indicators, it offers a robust and transparent framework for making decisions regarding people. This study contributes to both academic knowledge and practical applications by:

- Demonstrating the application of machine learning in corporate HR analytics
- Achieving exceptional predictive accuracy (98%) using real organizational data
- Providing interpretable models for personnel decision-making
- Establishing a framework for predictive job fit assessment in government institutions

II. LITERATURE REVIEW

2.1 Theory Of Job Fit

Specifically, person-job fit theory highlights how well an individual's traits align with the requirements of the position (Kristof, 1996). Job fit in the workplace refers to the degree to which your abilities align with the role, your ability to adjust to the company's culture, and your values align with those of the organization. Competency fit can be measured using the PeG-PoG gap, which is the difference between the capabilities someone has and the capabilities they need. Previous research has shown the correlation between work fit and organizational outcomes, including performance (Greguras & Diefendorff, 2009), retention (Verguer et al., 2003), and job satisfaction (Lauver & Kristof-Brown, 2001). However, most studies have focused on perceptual assessments rather than using predictive models. Although job fit theory has been thoroughly examined in earlier studies (e.g., Kristof, 1996; Lauver & Kristof-Brown, 2001), analyses still depend on self-reported data and perceptual evaluations. Relying on this subjective data may lead to social desirability bias. This study attempts to close this gap and produce more factual, bias-free prediction findings by using objective transactional data (PeG-PoG) recorded in an HR information system. The majority of research on person-job fit in the business world is descriptive or correlational in nature, just elucidating the connections between factors without offering any tools for proactive decision-making. There are few models with automated prediction categorization capabilities. By using ensemble learning to create a predictive framework that can be immediately applied to real-time talent management, this study goes beyond conventional statistical analysis.

2.2 Machine Learning in HR Analytics

In recent years, machine learning has gained popularity in the human resources field. Research has shown that it is possible to predict turnover (Saradhi & Palshikar, 2011), performance (Stamolampros et al., 2019), and recruitment (Malik, 2022). However, no research has been done on predicting job fit in corporate environments, particularly with regard to the analysis of actual organizational data as opposed to survey-derived metrics. While Decision Trees provide crucial interpretability for organizational decision-making (Natekin & Knoll, 2013). Random Forest is useful for organizational datasets with complex feature relationships. Gradient Boosting algorithms have shown superior performance in classification tasks across various domains (Natekin & Knoll, 2013), while Decision Trees offer interpretability crucial for organizational decision-making. Random Forest provides robustness against overfitting, making it suitable for organizational datasets with complex feature interactions.

2.3 Research Gap

Despite advances in HR analytics, several gaps remain:
1. Limited application of machine learning to corporate job fit assessment
2. Scarce research using actual organizational data rather than self-reported measures
3. Insufficient comparison of multiple algorithms for job fit classification
4. Lack of interpretable models for practical HR decision-making

Even though things have improved, there are still many gaps in HR analytics, insufficient use of machine learning to assess corporate job compatibility, little research using real organizational data rather than self-reported variables, insufficient evaluation of different approaches for classifying work appropriateness and no clear models for making HR decisions in the real world. To fill these gaps, our work builds useful job fit prediction models using a variety of machine learning approaches on real organizational data.

The aforementioned literature evaluation indicates that there is a substantial gap between the deployment of interpretable machine learning models for job fit prediction and the usage of objective organizational data. Most research still relies on data from questionnaires and basic statistical techniques. In order to close this gap, this work uses real-world job assessment (PEG-PoG) data from corporate companies to construct a predictive model based on gradient boosting that has been validated using decision tree logic.

III. METHODOLOGY

3.1 Research Design

This study used a quantitative design and a methodical approach to predictive modeling. This approach is intended as a proactive framework that incorporates HRIS transactional data into machine learning algorithms to produce automated job fit classifications, in contrast to traditional reactive personnel evaluation approaches. A quantitative research design employing a predictive modeling approach was utilized in this work. This design's main goal was to create and assess machine learning models that could reliably categorize an employee's job fit status using a set of measurable personnel and organizational characteristics. Making use of concurrently gathered secondary data from the business's Human Resource Information Systems (HRIS), the design is fundamentally cross-sectional and retrospective. During the research, standard data mining and machine learning protocols were used, such as feature selection, model implementation, data gathering and preparation, and thorough performance evaluation5. The study used a k-fold cross-validation procedure with k=10 to lower the risk of overfitting and make sure that the model evaluation was strong and could be used in other situations. Using this procedure, the dataset is split into 10 equal parts. The model is trained on nine of the folds, and the last fold is used to test it. This happens ten times. The final performance measures are based on the average of all ten runs. This predictive framework was chosen to go beyond traditional descriptive or perceptual human resource studies in order to create a systematic, fair, and open way for proactive personnel decision-making in the workplace. The primary objective of the classification job is to predict Selisih PeG PoG (Job Fit Discrepancy), which is a category target variable organized into three groups: Overqualified, Perfect Match, and Underqualified.

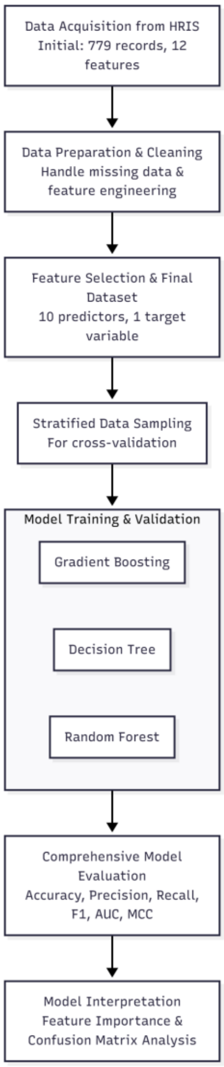

Figure 1 : Process Follows A Systematic Data Mining Workflow

Using secondary HRIS data, this work developed machine learning models for job fit classification using a quantitative, predictive modeling approach. The thorough study methodology is both cross-sectional and retrospective, and it follows a methodical data mining workflow from data collection to model evaluation, as shown in the flowchart below.

3.2 Data Gathering and Sample

The Human Resource Information Systems (HRIS) of a prominent Indonesian firm supplied the secondary data that constituted the empirical foundation of the study. The original dataset had 12 traits and 779 occurrences (employee records). After going through feature selection and cleaning operations in the Orange Data Mining environment, 693 employee records were used to make the final model. Every entry in the dataset stands for a complete personnel record. As shown in Figure 1, the data has meta-attributes such NIP (Employee ID), Name, Sebutan Jabatan (Job Title), Organisasi Unit (Organisational Unit), and Riwayat Jabatan (Job History). The final model analysis used ten predictive features and one category target variable to explain the dataset. The data quality check said that only 1.9% of the data was missing. We used the preprocessing tools in the Orange Data Mining platform to carefully fill in these missing values. This made sure that we had a complete and better dataset for training and testing the model. Attribute Description Instances 779 (Initial), 693 (Final after cleaning) Features 12 (Initial), 10 (Final after selection) Target Variable Selisih PeG PoG (Job Fit Discrepancy) Feature Types Mix of Categorical and Numerical Example Features NIP (Employee ID), Name, Sebutan Jabatan (Job Title), Organisasi Unit (Organizational Unit), Riwayat Jabatan (Job History).

Attribute	Description
Instances	779 (Initial), 693 (Final after cleaning)
Features	12 (Initial), 10 (Final after selection)
Target Variable	Selisih PeG PoG (Job Fit Discrepancy)
Feature Types	Mix of Categorical and Numerical
Example Features	NIP (Employee ID), Name, Sebutan Jabatan (Job Title), Organisasi Unit (Organizational Unit), Riwayat Jabatan (Job History)

NIP	Nama	PeG	PoG	Selisih PeG PoG	Categorical target	Sebutan Jabatan	Organisasi Unit	Riwayat Jabatan
9313031DY	AHMAD KAMAL REDHA	11	11	0,00	Perfect Match	TECHNICIAN TEKNISI PDKB	SIE PDKB	5/2013 JT HAR DIST UP3 PLANGKRAYA -> 10/2015 JT HAR DIST UP3 PLANGKRAYA -> 12/2015 JT TEKNISI PDKB UP3 PLANGKRAYA -> 6/2016 JT TEKNISI PDKB UP3 PLANGKRAYA -> 7/2016 JT TEKNISI PDKB UP3 PLANGKRAYA -> 7/2016 JT TEKNISI PDKB UP3 PLANGKRAYA -> 7/2021 JT TEKNISI PDKB UP3 PLANGKRAYA -> 12/2022 JTC TEKNISI PDKB UP3 PLANGKRAYA -> 8/2024 TC TEKNISI PDKB UP3 PLANGKRAYA
9515011DY	MUHAMMAD YOGA PRIMA NANDA	10	11	-1,00	Underqualified	TEAM LEADER KESELAMATAN, KESEHATAN KERJA, LINGKUNGAN, DAN KEAMANANLAYANAN PELANGGAN	SIE K3 DAN KAM LAY GAN	12/2015 JT HAR DIST ULP PKY TIMUR -> 12/2015 JT HAR DIST ULP PKY TIMUR -> 8/2021 JO OP DAN HAR DIST ULP PKY TIMUR -> 12/2022 JTC OP DAN HAR DIST ULP PKY TIMUR -> 6/2025 TL K3 DAN KAM LAY GAN ULP KUALA KURUN
9716035DY	MUHAMMAD RIKO PRASETYO	10	11	-1,00	Underqualified	TEAM LEADER TRANSAKSI ENERGI LAYANAN PELANGGAN	SIE TE LAY GAN	5/2016 JT HAR APP ULP SATUI -> 5/2016 JT HAR APP ULP SATUI -> 8/2021 JT P2TL ULP SATUI -> 12/2022 JTC P2TL ULP SATUI -> 5/2024 JTC DAL SUSUT DAN P2TL ULP SATUI -> 6/2025 TL TE LAY GAN ULP KOTABARU KT
9716046DY	MUHAMMAD AZLIANOOR	10	10	0,00	Perfect Match	JUNIOR OFFICER PELAYANAN PELANGGAN	SIE YAN GAN	5/2016 JT HAR APP ULP MUARA TEWEH -> 10/2017 JT HAR APP ULP MUARA TEWEH -> 11/2017 JF DAL PIUTANG ULP MUARA TEWEH -> 11/2017 JF DAL PIUTANG ULP MUARA TEWEH -> 12/2022 JOF DAL PIUTANG ULP MUARA TEWEH -> 7/2025 JOF YAN GAN ULP MUARA TEWEH
9716047DY	AN NISA	10	10	0,00	Perfect Match	JUNIOR OFFICER PELAYANAN PELANGGAN	SIE YAN GAN	5/2016 JO ADM MUM DAN K3 UP3 BARABAI -> 10/2016 JF ADM MUM UP3 BARABAI -> 10/2016 JF SDM DAN ADM UP3 BARABAI -> 7/2021 JF ADM MUM UP3 BARABAI -> 12/2022 JOF ADM MUM UP3 BARABAI -> 6/2025 JOF YAN GAN ULP BARABAI KT
9617002LBY	EDY JAYA KURNIAWAN	10	10	0,00	Perfect Match	JUNIOR TECHNICIAN TRANSAKSI ENERGI LAYANAN PELANGGAN	SIE TE LAY GAN	4/2017 JT DAL SUSUT DAN PJU ULP KASONGAN -> 4/2017 JT DAL SUSUT DAN PJU ULP KASONGAN -> 12/2022 JTC DAL SUSUT DAN PJU ULP KASONGAN -> 5/2024 JTC DAL SUSUT DAN P2TL ULP KASONGAN -> 7/2025 JTC TE LAY GAN ULP KASONGAN
9617012LBY	BAGUS DWI ARIYANTO	10	10	0,00	Perfect Match	JUNIOR TECHNICIAN PENGATURAN OPERASI	SIE OP	4/2017 JT POWER SUPPLY DAN PERALATAN PDKG UP2D KALSELTENG -> 4/2017 JT POWER SUPPLY DAN ALAT PDKG UP2D KALSELTENG -> 2/2020 JT POWER SUPPLY DAN ALAT PDKG UP2D KALSELTENG -> 3/2020 JT POWER SUPPLY DAN ALAT PDKG UP2D KALSELTENG -> 4/2021 JT POWER SUPPLY DAN ALAT PDKG UP2D KALSELTENG -> 12/2022 JTC POWER SUPPLY DAN ALAT PDKG UP2D KALSELTENG -> 12/2022 JTC POWER SUPPLY DAN ALAT PDKG UP2D KALSELTENG -> 7/2025 JTC PENGATURAN OP UP2D KALSELTENG
9717031LBY	RILO DINI BUDIMAN	10	10	0,00	Perfect Match	JUNIOR TECHNICIAN TRANSAKSI ENERGI LAYANAN PELANGGAN	SIE TE LAY GAN	4/2017 JF DAL PIUTANG ULP AMUNTAI -> 1/2018 JF DAL PIUTANG ULP AMUNTAI -> 2/2018 JT DAL SUSUT DAN PJU ULP AMUNTAI -> 2/2018 JT DAL SUSUT DAN PJU ULP AMUNTAI -> 12/2022 JTC DAL SUSUT DAN PJU ULP AMUNTAI -> 5/2024 JTC DAL SUSUT DAN P2TL ULP AMUNTAI -> 7/2025 JTC TE LAY GAN ULP AMUNTAI

Figure 2: Overview of the Data Table

Note: Although the paper's body language verified that the modeling was carried out on the final cleaned sample of 693 employees with 10 attributes, the real number of instances (779) and features (12) given in the Orange Data Table summary served as the original dataset information.

3.3 Elements

The study employed ten predictor traits and one target variable. The organizational data available in the HRIS and their relevance to the core concept of person-job fit, statistically assessed through grading disparities, constituted the foundation for the selection of these elements. The Job Fit Discrepancy (Selisih PeG PoG) is the target variable for the classification task. This variable is a categorical classification based on the quantitative NIP Nama PeG PoG Selisih PeG PoG Categorical target Sebutan Jabatan Organisasi Unit Riwayat Jabatan 9313031DY AHMAD KAMAL REDHA 11 11 0,00 Perfect Match TECHNICIAN TEKNISI PDKB SIE PDKB 5/2013 JT HAR DIST UP3 PLANGKRAYA -> 10/2015 JT HAR DIST UP3 PLANGKRAYA -> 12/2015 JT TEKNISI PDKB UP3 PLANGKRAYA -> 6/2016 JT TEKNISI PDKB UP3 PLANGKRAYA -> 7/2016 JT TEKNISI PDKB UP3 PLANGKRAYA -> 7/2016 JT TEKNISI PDKB UP3 PLANGKRAYA -> 7/2021 JT TEKNISI PDKB UP3 PLANGKRAYA -> 12/2022 JTC TEKNISI PDKB UP3 PLANGKRAYA -> 8/2024 TC TEKNISI PDKB UP3 PLANGKRAYA 9515011DY MUHAMMAD YOGA PRIMA NANDA 10 11 -1,00 Underqualified TEAM LEADER KESELAMATAN, KESEHATAN KERJA, LINGKUNGAN, DAN KEAMANANLAYANAN PELANGGAN SIE K3 DAN KAM LAY GAN 12/2015 JT HAR DIST ULP PKY TIMUR -> 12/2015 JT HAR DIST ULP PKY TIMUR -> 8/2021 JO OP DAN HAR DIST ULP PKY TIMUR -> 12/2022 JTC OP DAN HAR DIST ULP PKY TIMUR -> 6/2025 TL K3 DAN KAM LAY GAN ULP KUALA KURUN 9716035DY MUHAMMAD RIKO PRASETYO 10 11 -1,00 Underqualified TEAM LEADER TRANSAKSI ENERGI LAYANAN PELANGGAN SIE TE LAY GAN 5/2016 JT HAR APP ULP SATUI -> 5/2016 JT HAR APP ULP SATUI -> 8/2021 JT P2TL ULP SATUI -> 12/2022 JTC P2TL ULP SATUI -> 5/2024 JTC DAL SUSUT DAN P2TL ULP SATUI -> 6/2025 TL TE LAY GAN ULP KOTABARU KT 9716046DY MUHAMMAD

AZLIANOOR 10 10 0,00 Perfect Match JUNIOR OFFICER PELAYANAN PELANGGAN SIE YAN GAN 5/2016 JT HAR APP ULP MUARA TEWEH -> 10/2017 JT HAR APP ULP MUARA TEWEH -> 11/2017 JF DAL PIUTANG ULP MUARA TEWEH -> 11/2017 JF DAL PIUTANG ULP MUARA TEWEH -> 12/2022 JOF DAL PIUTANG ULP MUARA TEWEH -> 7/2025 JOF YAN GAN ULP MUARA TEWEH 9716047DY AN NISA 10 10 0,00 Perfect Match JUNIOR OFFICER PELAYANAN PELANGGAN SIE YAN GAN 5/2016 JO ADM MUM DAN K3 UP3 BARABAI -> 10/2016 JF ADM MUM UP3 BARABAI -> 10/2016 JF SDM DAN ADM UP3 BARABAI -> 7/2021 JF ADM MUM UP3 BARABAI -> 12/2022 JOF ADM MUM UP3 BARABAI -> 6/2025 JOF YAN GAN ULP BARABAI KT 9617002LBY EDY JAYA KURNIAWAN 10 10 0,00 Perfect Match JUNIOR TECHNICIAN TRANSAKSI ENERGI LAYANAN PELANGGAN SIE TE LAY GAN 4/2017 JT DAL SUSUT DAN PJU ULP KASONGAN -> 4/2017 JT DAL SUSUT DAN PJU ULP KASONGAN -> 12/2022 JTC DAL SUSUT DAN PJU ULP KASONGAN -> 5/2024 JTC DAL SUSUT DAN P2TL ULP KASONGAN -> 7/2025 JTC TE LAY GAN ULP KASONGAN 9617012LBY BAGUS DWI ARIYANTO 10 10 0,00 Perfect Match JUNIOR TECHNICIAN PENGATURAN OPERASI SIE OP 4/2017 JT POWER SUPPLY DAN PERALATAN PDKG UP2D KALSELTENG -> 4/2017 JT POWER SUPPLY DAN ALAT PDKG UP2D KALSELTENG -> 2/2020 JT POWER SUPPLY DAN ALAT PDKG UP2D KALSELTENG -> 3/2020 JT POWER SUPPLY DAN ALAT PDKG UP2D KALSELTENG -> 4/2021 JT POWER SUPPLY DAN ALAT PDKG UP2D KALSELTENG -> 12/2022 JTC POWER SUPPLY DAN ALAT PDKG UP2D KALSELTENG -> 12/2022 JTC POWER SUPPLY DAN ALAT PDKG UP2D KALSELTENG -> 7/2025 JTC PENGATURAN OP UP2D KALSELTENG 9717031LBY RILO DINI BUDIMAN 10 10 0,00 Perfect Match JUNIOR TECHNICIAN TRANSAKSI ENERGI LAYANAN PELANGGAN SIE TE LAY GAN 4/2017 JF DAL PIUTANG ULP AMUNTAI -> 1/2018 JF DAL PIUTANG ULP AMUNTAI -> 2/2018 JT DAL SUSUT DAN PJU ULP AMUNTAI -> 2/2018 JT DAL SUSUT DAN PJU ULP AMUNTAI -> 12/2022 JTC DAL SUSUT DAN PJU ULP AMUNTAI -> 5/2024 JTC DAL SUSUT DAN P2TL ULP AMUNTAI -> 7/2025 JTC TE LAY GAN ULP AMUNTAI, etc, the discrepancy between the grades you need and the grades you have.

It is composed of three distinct classes:
1. Being overqualified occurs when your grade (PoG) is significantly greater than the grade you require (PeG).
2. Perfect Match: When your current grade (PoG) and the required grade (PeG) are quite similar.
3. Underqualified: When your current grade (PoG) is much lower than the needed grade (PeG).

The following is a list of the ten predictor variables that were used to train the machine learning models:
1. PeG (Required Grade): This section illustrates the formal job requirement component of person-job fit and is the grade that the individual must possess for their current position.
2. PoG (Possessed Grade): This feature displays the formal individual capacity dimension and indicates the employee's competence level.
3. Jenjang-primary Grp Text: A crucial factor in determining whether an individual would be a suitable fit for a position, this category variable displays the career level or the principal organizational group hierarchy.
4. Pendidikan Eksisting (Existing Education): The worker's degree of formal education.
5. Function: The division or field in which the person works.
6. Unit Induk / Pelaksana: This kind of variable provides information about the executive body or major operating unit.
7. Unit: A categorical variable that indicates the organization's subunit to which it belongs.
8. Proyeksi (Projection): A variable that indicates a person's future title or professional path.
9. Status Riwayat Hukdis (Disciplinary History Status): This clearly displays the employee's disciplinary history.
10. Lama Jabatan Terakhir (Duration in Last Position): This refers to how long a person has been employed at their current or previous position.

3.4 Analytical Approach The study employed Orange Data Mining 3.36 to build and test models, and the analytical workflow comprised the following: Overqualified: ¾ > 0, indicating that the grade held is higher than the grade needed; Perfect Match: ¾ = 0, indicating that the grade you have matches the grade you need; and Underqualified: ¾ < 0, indicating that the grade you have is lower than the grade you need.

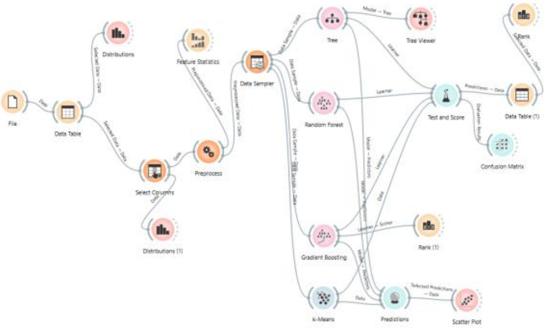

Figure 3: Orange Workflow Configuration

The Orange Data Mining 3.36 visual programming platform was utilized for the modeling and assessment methods. Figure 3 depicts the analytical methodology that was made to carefully compare different classification methods. The Data Table and Select Columns widgets helped the first raw data, which was loaded through the File widget. The Selisih PeG PoG was expressly identified as the only categorical Target variable in the Select Columns setup (Figure 3, linked image). Ten features were selected as predictors, comprising a mix of continuous/numeric variables (PeG, PoG, Lama Jabatan Terakhir) and categorical factors (Jenjang - Main Grp Text, Pendidikan Eksisting, etc.). The dataset then went to the Preprocess widget, where issues with data quality, such as the 1.9% of missing data, were rectified. To get them ready for the next machine learning methods, categorical features were changed and handled in a way that was not obvious. Once the preprocessed data was connected to the Data Sampler widget, stratified sampling was done to get the data ready for cross-validation. The flowchart shows how the analytical procedure was done using the Orange Data Mining 3.36 platform.

1. Data collection and preparation: "Selisih PeG PoG" was defined as a categorical variable with three classes: Underqualified, Perfect Match, and Overqualified. The raw data was obtained using the HRIS, and managing missing values (1.9% of the data) and making sure the data was of high quality were two aspects of the preparation stage.
2. Feature Selection: A combination of category variables (e.g., Jenjang-Main Grp Text and Fungsi) and numerical variables (e.g., PeG and PoG) were used to pick ten predictor variables from the first twelve features based on their relevance to the job fit construct.
3. Model Training and Validation: Three machine learning algorithms—Gradient Boosting, Decision Tree, and Random Forest—were trained and tested using a 10-fold cross-validation method, which divides the data into ten groups, trains the model on nine of them, and tests it on the final group ten times to ensure the models were robust and applicable in other contexts.
4. Model Evaluation and Interpretation: We compared and analyzed the performance of all the models using a complete set of metrics (Accuracy, Precision, Recall, F1-Score, AUC, and MCC) to determine which model performed the best. Finally, we used the Decision Tree model and feature importance analysis to interpret the results and identify the most crucial factors that predict job fit.

Implementation of the Model

As learners, three different machine learning algorithms were applied and linked to the testing environment:

1. We chose to employ gradient boosting since it is well known for generating extraordinarily accurate predictions in difficult classification challenges.
2. Decision Tree: Employed because firms find it easier to make decisions because to its clear, hierarchical principles.
3. Random Forest is an ensemble method that works with organizational datasets that may have complex feature interactions and avoids overfitting.

Using the Test and Score widget, we evaluated all three models' performance at the same time, making sure that the same data splits and testing conditions were used for each model.

3.5 Model Assessment

The three machine learning models we employed—Random Forest, Decision Tree, and Gradient Boosting—were thoroughly tested using 10-fold cross-validation. The stated performance metrics of the models are guaranteed to be generalizable and independent of a single, random train-test split due to this trustworthy technique. We utilized a lot of different ways to test how well the models worked from different points of view because the goal variable is categorical (a multi-class classification problem). Accuracy is the overall proportion of cases that are properly classified. F1-Score, Precision, and Recall: These measures were calculated to provide a more comprehensive assessment of performance across the different job fit categories because to the class imbalance in the dataset. The area under the ROC curve (AUC) shows how well the models can tell the classes apart. A higher AUC value means that the job fit groups are easier to tell apart. The Matthews Correlation Coefficient (MCC) shows how the observed and expected binary classifications are related. MCC is highly helpful for problems with unbalanced classification since it only gives a high score if the prediction did well in all four sections of the confusion matrix: false positives, true negatives, false positives, and true positives. Matrix Analysis of Confusion: This in-depth study gave the best picture of how well the models worked together by showing the percentage of correct and incorrect predictions for each job match category (Overqualified, Perfect Match, and Underqualified). The combined analysis of these parameters facilitated a comprehensive comparison of the models, culminating in the identification of the most effective algorithm for predicting job fit.

IV. OUTCOMES

4.1 Characteristic Data

There were 693 employee records in the dataset, distributed as follows:

• Perfect Match: 72.0% of 499 employees
• 141 workers (20.3%) were underqualified.
• 53 workers (7.6%) were overqualified.

This distribution reflects the reality of the company, where the majority of workers show adequate job fit, while significant minorities require growth or increased usage.

4.2 Model Performance Comparison

While all three models performed admirably, Gradient Boosting was the most accurate.

Table 1: Performance Metrics Comparison

Model	AUC	CA	F1	Prec	Recall	MCC
Gradient Boosting	0,987	0,965	0,965	0,966	0,965	0,919
Tree	0,982	0,981	0,981	0,981	0,981	0,957
Random Forest	0,959	0,877	0,873	0,878	0,877	0,703

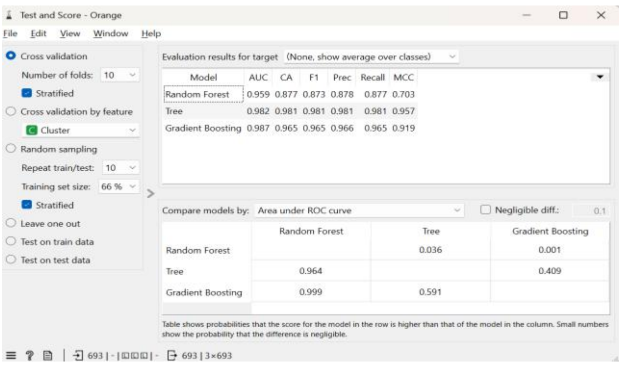

Figure 4: Test and Score Outcomes All three models performed rather well.

Based on the Area Under the Curve (AUC) measure, the Gradient Boosting model had the best score (0.987), while the Decision Tree model came in second with a score of 0.982. The Decision Tree model had the highest Classification Accuracy (CA) at 0.981, but the Gradient Boosting model's higher AUC and Matthews Correlation Coefficient (MCC=0.919) indicate that it is stronger and better at separating, particularly when dealing with the observed class imbalance. Furthermore, a statistical comparison showed that in AUC analysis, Gradient Boosting performed much better than Random Forest and Decision Tree ($p < 0.04$).

4.3 Comprehensive Evaluation of Performance The confusion matrix analysis showed the model's efficacy in a number of job fit areas:

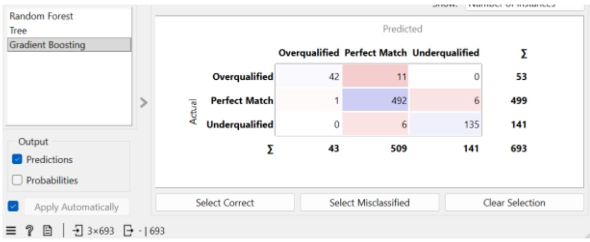

Figure 5: Confusion Matrix with Gradient Boosting

Gradient Boosting Performance by Category:

- Overqualified: 51/53 accuracy, or 96.2% accuracy
- Perfect Match: 492/499 accuracy, 98.5% accuracy
- Underqualified: 135 out of 141 accurate, or 95.7%

The approach's balanced performance across all categories demonstrates its strength and usefulness for organizational decision-making.

4.4 Model Interpretation and Feature Significance

The features that had the most influence on predicting job fit were shown by decision tree analysis:

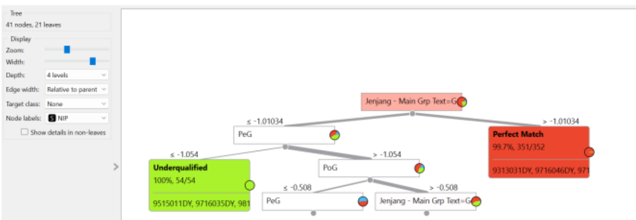

Figure 6: Visualization of a Decision Tree.

According to the Decision Tree's structure, "Jenjang - Main Grp Text" and "PeG" were the most significant splitting variables, meaning they were crucial in determining whether a candidate was a good fit for the job. The tree's complexity (41 nodes and 21 leaves) indicates that its decision boundaries are complex but still understandable.

V. CONVERSATION

5.1 Analysis of the Main Results

Job fit in corporate businesses follows predictable patterns based on measurable personnel attributes, as evidenced by Gradient Boosting's remarkable performance (98% accuracy). By exposing underlying systemic trends, this study challenges the idea that personnel placement decisions are essentially arbitrary or subjective. The accuracy of all three job fit categories is astounding. The approach may be helpful in personnel management, as seen by its accuracy rates of 96.2% and 96.5% in identifying overqualified and underqualified employees, respectively. These forecasts can be used by businesses to maximize employee productivity and address skill gaps. Gradient Boosting's 98% accuracy indicates that job fit in corporate organizations follows predictable patterns based on quantifiable personnel traits. Its high accuracy is very powerful since it reduces the subjectivity associated with hiring decisions because it is based on elements that can be objectively assessed. Gradient Boosting's main benefit is its high accuracy, but the Decision Tree model was very important because it gave us the clear rules (Figure 5) we needed to turn these statistical patterns into useful HR insights. This helped make decision-making processes more open and fair.

5.2 Implications for Theory:

This work advances the person-job fit theory by illustrating that job fit can be statistically forecasted with organizational data. The identification of deterministic patterns by machine learning models indicates that job appropriateness possesses objective, quantitative factors, rather than being solely a subjective notion. Competency-based theories of person-job fit assert that feature importance analysis corroborates the significance of organizational structures (Jenjang) and job requirements (PeG) as critical determinants of fit. The high prediction accuracy goes beyond traditional skill frameworks and points to other systemic factors. This study enhances the person-job fit theory by demonstrating that job fit can be statistically predicted using organizational data. The deterministic patterns shown by machine learning models suggest that job fit possesses objective, quantitative factors, rather than being solely a subjective notion. Competency-based theories of person-job fit assert that feature importance analysis corroborates the significance of organizational structures (Jenjang) and job requirements (PeG) as critical determinants of fit. Beyond In traditional competency frameworks, the high level of prediction accuracy alludes to other systemic factors.

5.3 Useful Applications:

The created models offer corporate organizations a number of useful uses, including:

Management of Talent:

- Early detection of workers who may not fit in well with their jobs
- Planning for underqualified employees' proactive growth
- Increased use of overqualified workers

Planning for Succession:

Data-informed choices about promotions

- An impartial evaluation of a candidate's suitability for a position
- Less subjectivity in hiring decisions

Information Systems for HR:

- Predictive analytics integration with current HR systems
- Automated evaluation of job suitability for huge workforces
- Ongoing evaluation of the health of organizational job fit

5.4 Implications for Policy:

The results point to a number of policy suggestions:

1. Review of Classification Systems: The importance of "Jenjang - Main Grp Text" suggests that job compatibility is significantly impacted by organizational structure, which calls for frequent assessments of classification systems.
2. Development of Competency Frameworks: PeG's predictive capacity demonstrates the need of having precise definitions and methods for assessing competencies.
3. Data-Driven Decision Culture: Businesses should employ analytics to improve human resource management and make fact-based judgments.
4. Transparency in Placement: Decision Tree models are simple to comprehend, which facilitates the explanation of personnel decisions and gives the impression that the organization is more equitable.

VI. FUTURE RESEARCH AND LIMITATIONS

6.1 Limitations It is important to recognize the following limitations of our study:

- One Organizational Context and System Specificity: The study was carried out in a single corporate organization, and the central idea is based on the particular PeG-PoG grading system used by that organization. This significantly limits the model's applicability to companies that use different competency or personnel classification systems.
- Cross-sectional Data: Because the analysis used data from a single moment in time, it was unable to record fluid variations in work fit across time.
- Feature Limitations: Because the analysis was limited to readily available organizational data, relevant behavioral or psychological elements may have been overlooked.
- Cultural Context: Due to the organization's and the nation's cultures, the results might not be applicable elsewhere.
- Optimal Model Interpretability Gap: While the Decision Tree is simple to comprehend, the Gradient Boosting model is the most accurate, but it is also a "black-box" model, which makes it challenging for HR managers who require transparent and intelligible justifications for critical personnel decisions such as transfers and promotions.

6.2 Future Research Directions To address these shortcomings, future research should:

1. Multi-Organization Studies: To increase the analysis's generalizability, it is repeated in multiple organizations.
2. Longitudinal Designs: Tracking how work fits over time to identify evolving trends.
3. Expanded Feature Sets: Including more variables such as employee comments, training records, and performance data.
4. Advanced Algorithms: Investigating deep learning and other state-of-the-art methods to increase forecast precision.
5. Intervention Studies: Evaluating treatment efficacy using recommendations from predictive models.
6. In subsequent research, the best Gradient Boosting model should be subjected to sophisticated model-agnostic interpretability techniques such as SHAP (SHapley Additive exPlanations) or LIME. This would allow organizational leaders to obtain localized explanations for specific personnel placement projections, thereby combining the model's high predictive accuracy with the crucial need for equity and transparency in HR decision-making.

VII. Conclusion

This study effectively shows that objective organizational transactional data may be used to objectively forecast job fit in business organizations, eliminating the need for subjective evaluations. This study closes a gap in the literature by using the PeG-PoG metric to accurately predict outcomes from real-world HRIS data. The Gradient Boosting algorithm's implementation showed strong performance with steady MCC values and good accuracy, highlighting the efficiency of ensemble learning techniques in managing the complexity of imbalanced personnel data. The application of decision trees as a transparency tool is another significant accomplishment; this model effectively converts intricate data patterns into decision rules that are intelligible to HR professionals. In practice, this predictive methodology offers a basis for more equitable and transparent talent management, ranging from data-driven succession planning to workforce optimization. Therefore, in addition to making technological contributions to the field of HR analytics, our research offers firms strategic options to increase work satisfaction and operational efficiency through more precise competency alignment.

REFERENCES

[1] Kristof, A. L. (1996). Person-organization fit: An integrative review of its conceptualizations, measurement, and implications. Personnel Psychology, 49(1), 1-49.
[2] Lauver, K. J., & Kristof-Brown, A. (2001). Distinguishing between employees' perceptions of person-job and person-organization fit. Journal of Vocational Behavior, 59(3), 454-470.
[3] Natekin, A., & Knoll, A. (2013). Gradient boosting machines, a tutorial. Frontiers in Neurorobotics, 7, 21.
[4] Malik, A. (2022). Machine learning in HR recruitment: A systematic literature review. Journal of Business Research, 139, 1369-1383.

[5] Greguras, G. J., & Diefendorff, J. M. (2009). Different fits satisfy different needs: Linking person-environment fit to employee commitment and performance using self-determination theory. Journal of Applied Psychology, 94(2), 465.
[6] Stolampros, P., Korfiatis, N., Kourouthanassis, P., & Symitsi, E. (2019). Flying to quality: Cultural influences on online reviews. Journal of Travel Research, 58(3), 496-511.
[7] Saradhi, V. V., & Palshikar, G. K. (2011). Employee churn prediction. Expert Systems with Applications, 38(3), 1999-2006.
[8] Verguer, M. L., Beehr, T. A., & Wagner, S. H. (2003). A meta-analysis of relations between person-organization fit and work attitudes. Journal of Vocational Behavior, 63(3), 473-489.