

# A Comparative Study of Naive Bayes and Decision Tree Classifiers for Forecasting Attrition Risk Among AI Professionals Based on Burnout Indicators

Aprian Java Dillah<sup>1\*</sup>, Budi Santoso<sup>2</sup>

<sup>1,2</sup> Informatika, Bina Sarana Informatika University  
[javadillah@gmail.com](mailto:javadillah@gmail.com)<sup>1\*</sup>, [budi.bis@bsi.ac.id](mailto:budi.bis@bsi.ac.id)<sup>2</sup>

## Abstract

Employee turnover in artificial intelligence (AI)-related roles is closely linked to burnout, yet many organizations still lack a systematic way to flag at-risk staff before they resign. This study addresses that gap by building and comparing two supervised classifiers, Naive Bayes and Decision Tree, to forecast attrition risk from workplace burnout signals. A publicly available Kaggle dataset comprising 1,500 employee records across 19 attributes was cleaned through mean/mode imputation and narrowed down via feature selection before being split into training and testing subsets. Using Orange Data Mining together with 10-fold cross-validation, the two models were benchmarked twice: once on the full feature set and once after irrelevant attributes were removed. Prior to feature selection, Decision Tree slightly led with 78.8% accuracy against Naive Bayes' 77.6%; after selection, the ranking reversed, with Naive Bayes reaching 80.2% accuracy versus 80.1% for Decision Tree. The comparison suggests that neither algorithm is universally superior performance instead hinges on how the underlying features are curated, a finding that should inform how HR analytics teams design early-warning systems for burnout-driven attrition.

**Keywords:** Attrition; Burnout; Classification; Decision Tree; Machine Learning; Naive Bayes

## 1. Introduction

Losing a skilled employee rarely costs an organization only the position itself. Replacement expenses alone can exceed a full year of that employee's salary once recruitment, onboarding, and lost momentum are accounted for, which is why voluntary turnover has become a standing concern for HR departments rather than an occasional inconvenience [1]. The problem is not merely financial: teams left short-staffed after a departure often see productivity dip and revenue opportunities slip away in the interim.

A workforce that stays is, in effect, a workforce that trusts its employer. Organizations that succeed at retention tend to be the ones that keep working conditions comfortable, keep employees engaged, and keep job satisfaction high enough that people choose to remain rather than look elsewhere. This is precisely why the IT sector where skilled professionals are both scarce and mobile has become a testing ground for predictive retention analytics: by mining historical indicators such as satisfaction scores and engagement levels, machine learning models can flag employees who are quietly drifting toward resignation, well before an exit interview becomes necessary [2].

The rise of AI-driven tools inside the workplace complicates this picture further. On one hand, these tools promise faster, more objective ways to spot early stress or fatigue signals than a manager's intuition or an annual survey ever could. On the other hand, the very technologies meant to ease workloads have, since the pandemic reshaped work into remote, hybrid, and freelance arrangements, also introduced new fatigue triggers: the disappearance of a clear line between office hours and personal time, fewer face-to-face check-ins with colleagues, and a creeping "always-on" expectation that leaves little room to disconnect. Studies on flexible work arrangements confirm this double edge gains in autonomy and productivity, offset by a harder time drawing boundaries between work and rest [3],[4].

Given this tension, the present study sets out to test whether Naive Bayes or Decision Tree classifies AI-sector attrition risk more reliably when the input is burnout-related data, and whether trimming the feature set down to only the most informative attributes changes that verdict.

## 2. Literature Review

### 2.1. Conceptual Background

At its core, data mining is less about storing information and more about excavating it sifting through large repositories of records to surface relationships that would otherwise stay buried [5]. Burnout, in turn, is what Maslach's framework describes as a three-sided response to chronic occupational stress: emotional depletion, a cynical detachment from one's work (depersonalization), and a shrinking sense of personal accomplishment. Layer a digital workplace on top of that, and a fourth strain emerges "digital burnout," the exhaustion that comes from being perpetually reachable and juggling overlapping digital demands until work and life blur together.

Naive Bayes tackles classification through probability: it assumes, somewhat unrealistically but usefully, that features are independent of one another, and in exchange gets a model that trains fast even on sizable datasets. Decision Tree takes a different route, carving the data into branches based on whichever attribute yields the greatest information gain at each split, until a legible set of if-then rules emerges arguably its biggest selling point over a probabilistic "black box." Judging either model typically comes down to four numbers pulled from a confusion matrix accuracy, precision, recall, and F1-score while k-fold cross-validation guards against the risk that a single lucky (or unlucky) train-test split is driving the results. All experiments here were run inside Orange Data Mining, a node-based visual analytics environment [6], [7].

### 2.2. Related Work

Table 1 gathers a cross-section of earlier work on attrition- and burnout-related classification, spanning corporate HR datasets as well as academic-stress studies that share the same underlying algorithms.

**Table. 1 : Summary of Related Work**

| Ref. | Research Focus                                                        | Key Result                                                                        |
|------|-----------------------------------------------------------------------|-----------------------------------------------------------------------------------|
| [8]  | Pitting Naive Bayes against C4.5 on employee attrition records        | Naive Bayes edged ahead with 98.66% accuracy versus 96.26% for C4.5.              |
| [9]  | Building an early-warning attrition model from 1,010 employee records | The C4.5-based Decision Tree reached 92.9% accuracy in flagging resignations.     |
| [10] | Pairing clustering with classification to study IT-sector turnover    | K-Means combined with a classifier tracked turnover trends with good precision.   |
| [11] | Screening students for academic stress as a burnout precursor         | Naive Bayes topped the field at 93.26% accuracy; C4.5 trailed at 80.60%.          |
| [12] | Isolating stress predictors among graduating students                 | Interpersonal-relationship attributes emerged as the strongest stress predictor.  |
| [13] | Scoring burnout risk against the Maslach Burnout Inventory            | A C5.0 tree reached 99.25% training accuracy and 97% on unseen test data.         |
| [14] | Testing SMOTE against class imbalance in burnout classification       | SMOTE backfired on a small dataset, cutting Naive Bayes accuracy by 9.52%.        |
| [15] | Framing attrition prediction as an input to retention planning        | Logistic Regression, SVM, and Decision Tree each informed HR retention decisions. |

What distinguishes the present work from this body of literature is its narrower target population AI professionals specifically and its explicit before/after comparison of feature selection, a variable most of the studies above did not isolate as a standalone factor.

## 3. Research Method

The research followed a linear pipeline define the problem, gather data, prepare it, train competing models, then score them so that any performance gap between Naive Bayes and Decision Tree could be attributed to the algorithms themselves rather than to inconsistent handling of the data.

### 3.1. Dataset

Records were sourced from a Kaggle dataset describing 1,500 AI-sector employees (roles such as data scientists, ML engineers, and AI developers) across 19 attributes touching on job satisfaction, burnout intensity, apprehension about being replaced by AI, the share of tasks already automated, and daily hours spent working alongside AI tools. The label to be predicted, attrition risk, falls into three tiers: High, Medium, and Low.

### 3.2. Preprocessing and Feature Selection

Missing entries were filled in through Orange's Preprocess widget numeric gaps with the column mean, categorical gaps with the most frequent value leaving all 1,500 records intact rather than discarding incomplete rows. The attribute list was then pruned using the Select Columns widget: of the original 19 variables, five were kept as predictors (hours\_with\_ai\_assistance\_daily, ai\_replaces\_my\_tasks\_pct,

burnout\_score, job\_satisfaction\_1\_5, and fear\_of\_ai\_replacement), the rest were dropped, and employee\_id was demoted to a meta field used purely for record identification.

### 3.3. Data splitting

An 80:20 split set aside 1,200 records for training and 300 for testing, giving each model enough examples to learn from while keeping a large enough holdout sample to judge how well it generalizes to cases it has not seen.

### 3.4. Modeling and evaluation

Both classifiers were assembled inside Orange Data Mining following the same chain of widgets File, Select Columns, Preprocess, then either Naive Bayes or Tree, followed by Test and Score and Confusion Matrix so that the only variable changing between runs was the algorithm itself, as illustrated in Fig. 1. Ten-fold cross-validation supplied four comparison metrics (Classification Accuracy, F1-score, Precision, and Recall), captured twice: once before feature selection and once after, to isolate its effect on each model.

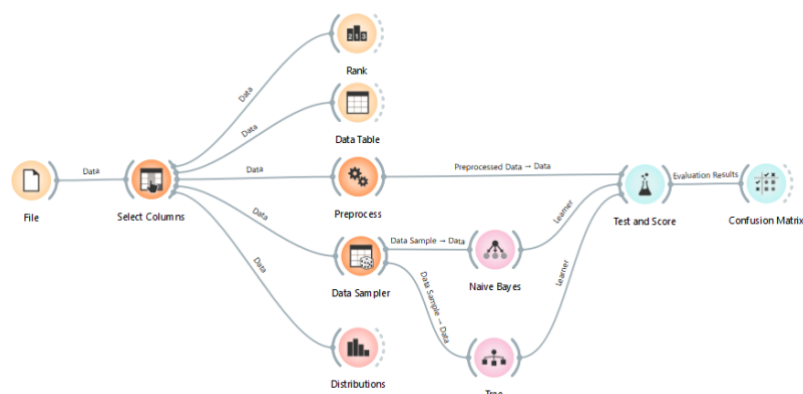


Fig. 1: workflow built in Orange Data Mining

Data-processing from data input through preprocessing, feature selection, model training (Naive Bayes / Decision Tree), and evaluation via Test and Score and Confusion Matrix

## 4. Results and discussion

The attrition\_risk label is anything but evenly spread across the 1,500 records. Low-risk cases make up the bulk of the data at roughly 700–725 instances, Medium-risk trails behind at around 600–650, and High-risk cases are scarce only about 75–100 records. That scarcity matters: with so few High-risk examples to learn from, both models start at a disadvantage when it comes to recognizing that class specifically, a caveat worth keeping in mind when reading the accuracy figures below.

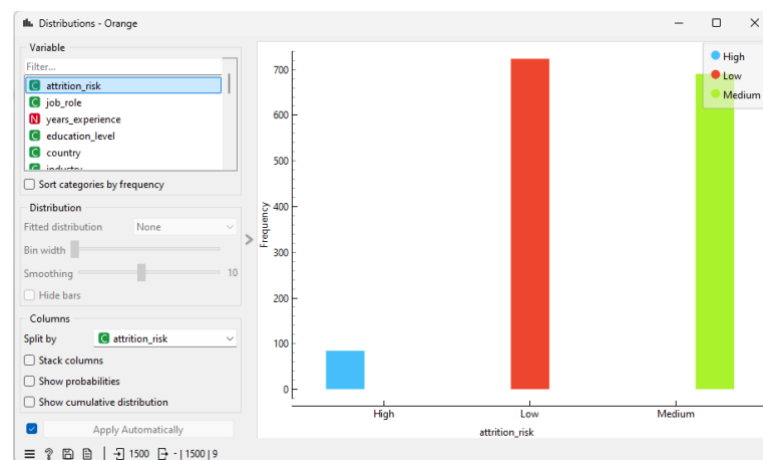


Fig. 2: Distribution Orange

of the attrition\_risk target class across the 1,500-record dataset, showing an imbalance in favor of the Low-risk category.

## 4.2. Performance before feature selection

Table 2 reports how each algorithm performed when trained on the full, unfiltered set of 19 attributes.

**Table 2:** Model Performance Before Feature Selection

| Algorithm     | CA    | F1    | Precision | Recall |
|---------------|-------|-------|-----------|--------|
| Naive Bayes   | 0.776 | 0.775 | 0.774     | 0.776  |
| Decision Tree | 0.788 | 0.787 | 0.786     | 0.788  |

At this stage Decision Tree held a small lead over Naive Bayes on every metric. Looking inside the confusion matrices, both models handled the Low class comfortably but stumbled on High, frequently mistaking it for Medium an early sign that some of the 19 raw attributes may be blurring the line between those two categories rather than helping separate them.

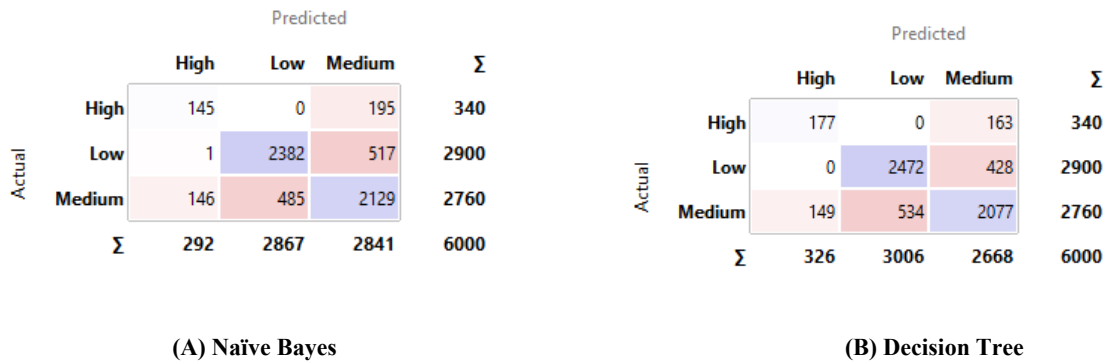


**Fig. 3:** Confusion matrices of the Naive Bayes and Decision Tree classifiers after feature selection. Decision Tree shows a marked improvement on the High class, while Naive Bayes' gains are concentrated in the Low and Medium classes.

## 4.3 Performance after feature selection

Table 3 shows the same two models retrained on just the five attributes retained after feature selection.

**Table 1.** Model Performance After Feature Selection

| Algorithm     | CA    | F1    | Precision | Recall |
|---------------|-------|-------|-----------|--------|
| Naive Bayes   | 0.802 | 0.797 | 0.801     | 0.802  |
| Decision Tree | 0.801 | 0.801 | 0.802     | 0.801  |

Trimming the feature set paid off for both algorithms Naive Bayes climbed from 77.6% to 80.2% accuracy, Decision Tree from 78.8% to 80.1%. But the confusion matrices in Fig. 4 reveal that this aggregate gain was not shared evenly across classes. Decision Tree's correct predictions on the High class jumped from 177 to 222 out of 340, implying that once the noisier attributes were removed, the remaining five features gave it a clearer signal for spotting high-risk cases specifically. Naive Bayes moved in the opposite direction on that same class its correct High-class predictions actually fell from 145 to 116 even as its correct Low-class predictions rose from 2,382 to 2,457 and its Medium-class predictions rose from 2,129 to 2,237. In other words, Naive Bayes' overall accuracy gain after feature selection was driven almost entirely by better performance on the majority classes, while Decision Tree was the algorithm that actually got better at catching the rare, high-risk employees the model exists to flag in the first place.

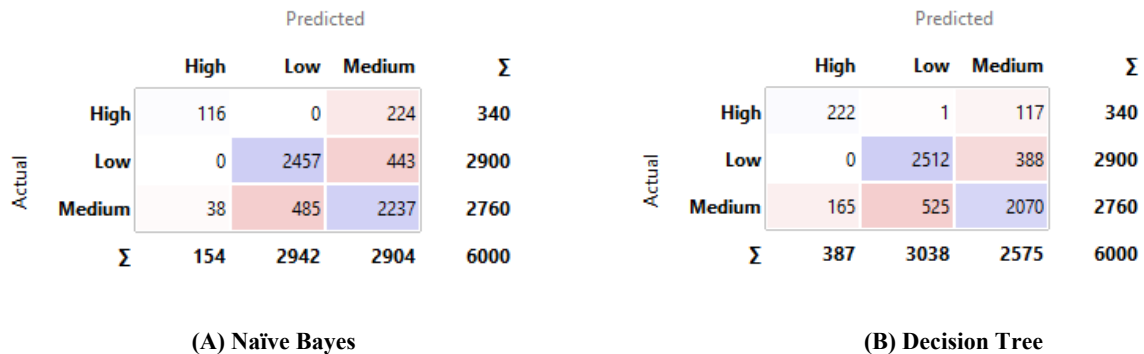


**Fig. 4:** Confusion matrices of the Naive Bayes and Decision Tree classifiers after feature selection. Decision Tree shows a marked improvement on the High class, while Naive Bayes' gains are concentrated in the Low and Medium classes.

#### 4.4 Comparison of model performance

Table 3 shows the same two models retrained on just the five attributes retained after feature selection.

**Table 3:** Comparison of Model Performance Before and After Feature Selection

| Algorithm     | CA (Before) | CA (After) | F1 (Before) | F1 (After) |
|---------------|-------------|------------|-------------|------------|
| Naive Bayes   | 0.776       | 0.802      | 0.775       | 0.797      |
| Decision Tree | 0.788       | 0.801      | 0.787       | 0.801      |

Put side by side, the two rounds of testing tell a story of a lead that changes hands: Decision Tree starts ahead, but Naive Bayes overtakes it once the feature set is refined, closing at 80.2% accuracy and recall against Decision Tree's 80.1%. The margin is thin enough that either model could reasonably be deployed, but Naive Bayes carries the added appeal of being cheaper to train and simpler to maintain, making it the more pragmatic pick for this particular dataset without discounting Decision Tree's advantage in producing rules a non-technical HR team could actually read and act on.

#### 5. Conclusion

Bringing the findings together, burnout-related signals how exhausted an employee reports feeling, how satisfied they are with their job, how threatened they feel by AI automation, how much of their work AI has already absorbed, and how many hours they log alongside AI tools daily carry enough signal to sort AI-sector employees into High, Medium, and Low attrition-risk bands. Naive Bayes moved from 77.6% to 80.2% accuracy once the feature set was trimmed, while Decision Tree moved from 78.8% to 80.1%; the post-selection edge belongs to Naive Bayes, though only narrowly. The larger takeaway is less about crowning a winning algorithm and more about the value of feature selection itself, which lifted both models regardless of which one is ultimately chosen. Later studies could push this further by testing on larger and more varied datasets, applying resampling techniques such as SMOTE to correct the High class scarcity noted here, and bringing additional algorithms Random Forest, SVM, or KNN into the comparison.

#### Acknowledgement

The author is grateful to the thesis supervisor and to Universitas Bina Sarana Informatika for their guidance and support throughout this research.

#### References

- [1] R. B. Lestari, D. L. Diana, F. Psikologi, and U. Persada, "Pengaruh Burnout & Job Satisfaction Terhadap Turnover Tenaga Profesional Teknologi Informasi PT 'X' & Subsidiary," vol. 3, no. 1, pp. 85–95, 2023.
- [2] A. R. B. Jamroni, W. Hadikristanto, and M. Fatchan, "Analisis Faktor dan Prediksi Atrisi untuk Optimalisasi Retensi Karyawan Menggunakan Machine Learning," vol. 7, no. 3, 2025, doi: 10.32877/bt.v7i3.2301.
- [3] J. Sains *et al.*, "Prediksi dan Pencegahan Risiko Burnout pada Pekerja Fleksibel Menggunakan Algoritma Random Forest," vol. 5, no. 1, pp. 21–30, 2025, doi: 10.47065/jussi.v5i1.8937.
- [4] R. L. Pratiwi *et al.*, "2,3,4,5," pp. 1132–1138, 2025.
- [5] F. Widiyanto, "Analisis Sentimen Komentar Youtube tentang Konflik Iran-Israel Menggunakan Orange Data Mining," *Sains Data J. Stud. Mat. dan Teknol.*, vol. 3, no. 2, pp. 81–88, 2025, doi: 10.52620/sainsdata.v3i2.278.
- [6] J. Manajemen and S. Bisnis, "ERA BARU DUNIA KERJA : BAGAIMANA AI MENGGESER PEKERJAAN RUTIN DAN MENCIPTAKAN PELUANG BARU," vol. 1, no. 1, pp. 44–51, 2025.
- [7] D. A. Punkastyo, F. Septian, and A. Syaripudin, "Implementasi Data Mining Menggunakan Algoritma Naïve Bayes Untuk Prediksi Kelulusan Siswa," *J. Syst. Comput. Eng.*, vol. 5, no. 1, pp. 24–35, 2024, doi: 10.61628/jsce.v5i1.1073.
- [8] F. Daewana, "Perbandingan Algoritma Naïve Bayes Dan C4 . 5 Dalam Prediksi Keputusan Karyawan Untuk Meninggalkan Perusahaan," vol. 5, pp. 1712–1726, 2025.
- [9] M. A. C, T. H. Lamramot, A. I. Hadiana, and I. Santikarama, "Sistem Prediksi Awal Terhadap Atrisi Karyawan," vol. 1, pp. 18–24, 2022.
- [10] S. K. Setianto, D. Jatikusumo, P. Study, T. Informatika, F. I. Komputer, and U. M. Buana, "Jurnal Mantik," vol. 4, no. 3, pp. 1573–1581, 2020.
- [11] N. Annisa, W. Pertiwi, and I. S. Wibisono, "Comparison of K-Nearest Neighbor , Naïve Bayes , and C4 . 5 Algorithms for Predicting Academic Stress Risk in Students Based on Psychological Survey Data," vol. 10, no. 1, pp. 698–706, 2026.
- [12] A. Trifani, "Penerapan Data Mining Klasifikasi C4 . 5 dalam Menentukan Tingkat Stres Mahasiswa Akhir," vol. 1, no. 2, 2022.
- [13] C. Algorithm and N. F. Rahmadani, "Prediction of Burnout Syndrome Risk in University Students Using the," vol. 13, no. 2, pp. 526–536, 2026, doi: 10.30865/jurikom.v13i2.9550.
- [14] M. A. F. F., "Klasifikasi Algoritma Naive Bayes Terhadap Data Imbalance," pp. 1–6, 2024.
- [15] R. Rianti and R. Andarsyah, "Prediksi Tingkat Atrisi Karyawan menggunakan Machine Learning," vol. 18, no. 1, pp. 39–52, 2024.