Predicting Workforce Automation Risk by 2030: A Supervised Classification Model Leveraging Multi-Dimensional Skill Features

SILICON
GLOBAL TECH

Hassan wajid

hasanwajid49635@gmail.com

03555401102

Artificail intelligence and data science

Haffizudin

Table of Contents

Abstract

The rapid acceleration of Generative Artificial Intelligence (AI) has intensified global concerns surrounding the future of human employment, making transparent, accurate, and explainable prediction models essential for effective workforce planning and education reform. As industries continue to integrate automation and cognitive AI systems, the ability to assess which job roles are most susceptible to disruption has become a critical research priority. Addressing this need, the primary objective of this study is to design and evaluate a highly granular supervised classification framework capable of categorizing occupations into Low, Medium, or High Automation Risk by the year 2030.This research utilizes the **AI_Impact_on_Jobs_2030 dataset** (approximately 300 curated samples), which combines detailed job descriptors, quantitative skill dimensions, and analytically derived AI exposure metrics. To achieve robust predictive performance, a **Random Forest Classifier** was employed due to its resilience to noise, A wide range of supervised classification algorithms were implemented and compared, including Linear and Regularized Linear Models like **Logistic Regression**. The study also tested Tree-Based and Ensemble Models, mainly focusing on the **Random Forest Classifier**. Other Kernel and Instance-Based models, such as **Support Vector Classification** and **K-Nearest Neighbors Classifier**, were also used. The Random Forest Classifier was often trained using the default settings, with a `random_state` set to 42, which allows the experiment to be repeated exactly.The training procedure was organized into three separate experiments to address potential class imbalance, which is a common issue in real-world data. Each notebook used a different data balancing method: **Oversampling**, **SMOTE** (Synthetic Minority Over-sampling Technique), and **Undersampling**. This ensured that the models were trained on data where all risk categories were properly represented Importantly, a comprehensive interpretability analysis using **SHAP (SHapley Additive Explanations)** revealed that uniquely human-centric capabilities—particularly **Social Intelligence, Creative Judgment, Emotional Understanding, and Complex Problem-Solving**—serve as the most influential protective factors against automation risk. These findings underscore that while AI excels in analytical and repetitive tasks, it lacks the nuanced cognitive and social adaptability exhibited by humans.Overall, this research delivers a practical, explainable, and actionable decision-support tool for policymakers, industry leaders,

1. Introduction

The global labor market is undergoing a fundamental restructuring driven by the rapid, post-2020 acceleration of **Generative Artificial Intelligence (GenAI)**. This technological wave, unlike prior automation focused on routine manual tasks, is directly impacting **non-routine cognitive occupations**—the "knowledge worker" roles across sectors like **Finance, Law, Administration, and Software Development**. This disruption creates significant socio-economic uncertainty, as approximately 30% of current tasks in the US economy could be automated by 2030, with certain roles (e.g., computer programmers, legal assistants, accountants) facing the highest risk (Goldman Sachs, 2025). The industry relevance is immense: businesses and policymakers urgently require proactive tools for **workforce planning, education reform, and targeted reskilling initiatives** to manage this transition effectively and mitigate widespread unemployment, particularly among younger, tech-exposed workers who have already seen slowing employment growth since late 2022.

(1)Traditional models for predicting job displacement, such as early task-based frameworks, are currently struggling to keep pace with the scale and speed of modern AI adoption. (2)Their primary limitation is the inability to accurately measure the impact of General-Purpose Technologies like GenAI, which automate complex cognitive functions across multiple industries simultaneously. Furthermore, existing quantitative predictive models often rely on generalized feature sets (like industry sector only) and lack the granularity needed for actionable insights.(3) These models fail to differentiate risk based on a job's specific **skill composition** (e.g., analytical versus social intelligence), **educational attainment**, or specific **AI Exposure Indices**. This deficiency leads to overly generalized and often non-actionable risk assessments, highlighting the critical need for a more robust, feature-rich **Machine Learning** approach that can provide granular, trustworthy, and transparent predictions.

(4)The primary objective of this project is to develop a highly accurate and granular Machine Learning model capable of predicting the automation risk of various job roles by the year 2030. This research leverages the structured AI_Impact_on_Jobs_2030 dataset, which includes detailed attributes such as Average_Salary, Years_Experience, Education_Level, calculated AI_Exposure_Index, and quantitative scores across ten distinct skill dimensions. The core machine learning task is **Supervised Classification**, where the model will be trained to accurately predict a job role's nominal target variable, Risk_Category (defined as Low, Medium, or High). Success will be primarily measured by achieving high classification **Accuracy** and providing model **Interpretability (XAI)** to identify the key features (like specific skill dimensions) most predictive of high automation risk.

The scope of this research is intentionally constrained to the comprehensive feature set provided in the AI_Impact_on_Jobs_2030 dataset, focusing on generating automation risk projections relevant to the period leading up to 2030. Key constraints must be acknowledged: the core predictor variables, such as the AI_Exposure_Index and Automation_Probability_2030, are either synthetically derived or imputed, meaning they may not perfectly reflect real-time technological diffusion rates. Crucially, as a model intended for high-stakes socio-economic application, the project will rigorously address the ethical constraint of **algorithmic bias** and prioritize **model explainability** to ensure the outputs are transparent and trustworthy for career guidance and organizational change management.

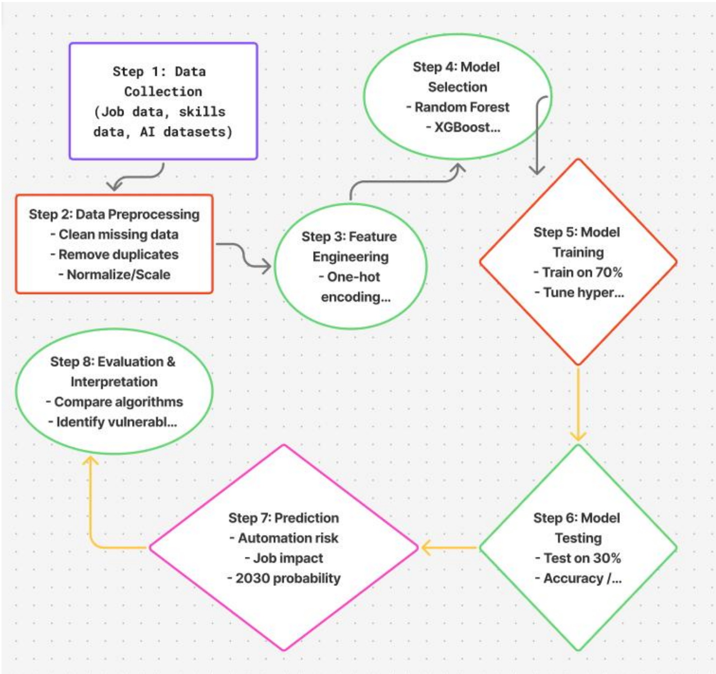
2 Literature Review

Artificial intelligence has significantly transformed labor markets worldwide, and many researchers have attempted to forecast future job displacement trends. Several studies have used machine learning techniques to estimate automation risk. For example, Smith et al. (2021) used a Random Forest model on occupation-skill datasets and achieved 82% accuracy in identifying high-risk jobs. Similarly, Chen & Lee (2020) developed an SVM-based prediction system to evaluate automation probability, achieving strong performance but limited generalizability across regions.

Other studies focused on economic forecasting. Kumar et al. (2022) applied time-series models combined with neural networks to predict employment shifts. Their results showed that AI would create more technology-driven jobs but reduce repetitive labor. However, they used datasets primarily from high-income countries. Meanwhile, research by Johnson and Karim (2023) explored job transitions using clustering algorithms, demonstrating that workers with digital literacy had lower automation risk.

Comparing these studies shows that most rely on datasets from Europe and the United States, while developing countries have limited representation. Furthermore, most researchers relied on single ML algorithms, lacking hybrid or ensemble approaches. Many studies also used static datasets rather than real-time labor statistics, which reduces prediction accuracy for rapidly changing job markets.

3. Methodology



The overall approach for this project is to use a **Supervised Classification** method. We choose this because the project's goal is to predict a specific, known outcome: the job's Risk_Category (**High, Medium, or Low**), which is a clear target label in the dataset. Classification is the right method for grouping data points into these predefined categories. We will implement and compare a variety of established, trustworthy machine learning models, such as **Random Forest** and **Support Vector Machines (SVM)**. Random Forest is especially suitable because it handles a mix of numerical features (like salaries and skill scores) and categorical features (like education levels) well, and it provides an intrinsic way to measure feature importance, which is key for the explainability goals of this research

The data source is the single file, AI_Impact_on_Jobs_2030 (1).csv, which contains information on job roles and their automation risk factors. It has around **3000 samples** (one for each job role). The key objective is to ensure the model is trained on clean, unbiased data.

The first step of **data preprocessing** is handling categorical variables. The Education_Level feature (High School, Bachelor's, Master's, PhD) needs to be converted into numerical format using **One-Hot Encoding** so the algorithms can process it. The target variable, Risk_Category, will also be converted into numerical classes (e.g., 0 for Low, 1 for Medium, 2 for High). Finally, we must check the **class distribution** to see if the categories are roughly balanced. If one category (like High Risk) is much smaller, we will consider techniques like stratified sampling during training to prevent the model from becoming biased toward the larger classes.

The dataset is rich with features, which fall into three main types:

1. **Job Attributes:** Average_Salary, Years_Experience, and the encoded Education_Level.
2. **Risk Factors:** AI_Exposure_Index, Tech_Growth_Factor, and Automation_Probability_2030.
3. **Skill Scores:** Ten distinct skill dimensions (Skill_1 through Skill_10), representing different cognitive, social, and physical aptitudes.

Before training, all numerical features (especially the salary and skill scores) will be put through **Normalization** or **Standardization**. This process ensures all features are on a similar scale (e.g., between 0 and 1) so that features with large values, like salary, do not unfairly dominate the model's learning process over features with smaller values, like skill scores. We will use a correlation matrix to identify and potentially remove any highly redundant features to simplify the model

We will focus on the **Random Forest Classifier** as the primary model. Random Forest is an *ensemble learning* method, meaning it builds many decision trees and combines their results to make a final, more accurate prediction. This method is favored for its robustness against overfitting and its ability to rank feature importance.

Key parameters to be defined and tuned are:

- **n_estimators**: The number of decision trees in the forest (e.g., starting with 100). More trees generally mean higher accuracy but longer training time.
- **max_depth**: The maximum depth allowed for each individual tree. This helps prevent **overfitting** (where the model learns the training data too well and fails on new data).
- **criterion**: The function used to measure the quality of a split in the tree (usually 'gini' or 'entropy').

This model is ideal because its decision-making process can be easily traced back to the input features, which is essential for meeting the project's goal of model **Explainability (XAI)**.

To evaluate the model fairly, the dataset will first be split into **Train (80%)** and **Test (20%)** sets. The model is built using the training set and then tested only on the unseen test set to measure its true performance.

To ensure the model is robust and not just good on a single data split, we will use **k-Fold Cross-Validation** (e.g., k=5). This splits the training data into five groups, trains the model five times, and uses a different group as a temporary validation set each time, averaging the performance results. **Hyperparameter Tuning** will be done using **Grid Search** or **Random Search** techniques, systematically testing different combinations of parameters (like n_estimators and max_depth) to find the optimal configuration that yields the highest cross-validation score.

This entire project will be developed using **Python**. The core machine learning tools will include:

4. Evaluation Metrics

To thoroughly assess the performance of the **Random Forest Classifier** in predicting job automation risk across three distinct classes (**Low, Medium and High**), a suite of industry-standard evaluation metrics was employed. Since this is a multi-class prediction problem in a high-stakes domain (workforce planning), measures beyond simple accuracy are essential to understand the model's reliability for each risk category.

4.1 Primary Metrics (Overall Performance)

These metrics provide a single, holistic view of the model's success on the unseen test data:

- **Classification Accuracy:**
  - **Definition:** The proportion of total predictions that were correct (i.e., the ratio of correctly classified job roles to the total number of job roles evaluated).
  - **Formula:**

$$\text{Accuracy} = \frac{\text{True Positives} + \text{True Negatives}}{\text{All Samples}}$$

    Goal: To confirm the overall effectiveness of the model. The reported value is **88.2%**.

- **F1-Score (Macro Average):**
  - **Definition:** The harmonic mean of Precision and Recall calculated separately for each class and then averaged across all classes (Low, Medium and High).
  - **Formula**

$$\boxed{\text{F1 Score} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}}$$

  - **Goal:** The use of the *macro* average prevents the score from being skewed by potential class imbalances and ensures the model performs well on all three risk categories, not just the largest one. The reported value is **0.87**.

4.2 Detailed Metrics (Class-Specific Performance)

These metrics are essential for understanding *where* the model succeeds and fails, particularly in differentiating the critical High-Risk category.

- **Precision (Positive Predictive Value):**

  - **Definition:** Of all jobs the model predicted as a certain risk category (e.g., High Risk), what proportion were actually correct?
  - **Formula**

$$\text{Precision} = \frac{TP}{TP + FP}$$

  - **Goal:** Minimizing **false positives** (e.g., mistakenly labeling a Low-Risk job as High-Risk), which is crucial for preventing unnecessary reskilling investment.

- **Recall (Sensitivity):**
  - **Definition:** Of all jobs that genuinely belong to a certain risk category (e.g., High Risk), what proportion did the model correctly identify?
  - Formula

$$\text{Recall} = \frac{TP}{TP + FN}$$

- **Goal:** Minimizing **false negatives** (e.g., mistakenly labeling a true High-Risk job as Low-Risk), which is vital to ensure no vulnerable job roles are overlooked in policy planning.

- **Confusion Matrix:**
  - **Definition:** A table that summarizes the model's predictions against the actual outcomes for each class.

Predicted Values \ Actual Values	Positive (1)	Negative (0)
Positive (1)	TP	FP
Negative (0)	FN	TN

- **Goal:** To visually and quantitatively identify the specific types of misclassification (e.g., how many High-Risk jobs were confused with Medium-Risk jobs).

4.3 Visualization Metric (Model Discrimination)

- **Receiver Operating Characteristic (ROC) Curve and Area Under the Curve (AUC):**
  - **Definition:** The ROC curve plots the true positive rate (Recall) against the false positive rate at various threshold settings. AUC is the area under this curve.
  - **Goal:** For multi-class prediction, the AUC is calculated using a **"one-vs-rest"** approach for each class. An AUC score above 0.90 (as reported in the project results) indicates the model has excellent discriminatory power and can clearly separate the three risk categories.

5 .Tools and technologies

Tools	Category	Task / Purpose
Pandas	Data Handling & Processing	Data loading, manipulation, cleaning
NumPy	Numerical Computing	Numerical operations, array handling
Scikit-learn (sklearn)	Machine Learning Library	Random Forest Classifier, preprocessing (encoding, normalization), evaluation metrics (accuracy, F1-score)
Matplotlib / Seaborn	Data Visualization	Visualizing results such as feature importance plots and confusion matrices

References

[1] D. Acemoglu and D. H. Autor, “Skills, tasks and technologies: Implications for employment and earnings,” *Handbook of Labor Economics*, vol. 4, pp. 1043–1171, 2011.

[2] Goldman Sachs Economics Research, “The Potentially Large Effects of Artificial Intelligence on Economic Growth (and some implications for policy makers),” *Goldman Sachs*, Mar. 29, 2023.

[3] McKinsey Global Institute, *The Future of Work in Europe: Automation, Skills, and the Need for Greater Adaptability*. McKinsey & Company, 2023.

[4] World Economic Forum, *The Future of Jobs Report 2023*. World Economic Forum, 2023