JSS MAHAVIDYAPEETA

JSS SCIENCE AND TECHNOLOGY UNIVERSITY

SRI JAYACHAMRAJENDRA COLLEGE OF ENGINEERING

MYSURU-570006

Subject Name: Machine Learning

Event –1: Project Presentation

Title: Global AI Job Trends, Market Analysis and Salary Prediction

Submitted by:

Roll no	USN	Name
55	02JST22UCS082	Purnaa L
04	01JST22UCS007	Aditi Katta
16	01JST22UCS067	Keerthana K V
30	01JST22UCS126	Raksha N Urs

Submitted to:

Prof Rakshitha R

Assistant Professor

Department of Computer Science and Engineering

JSS S&TU, SJCE Mysuru

Table Of Contents

1. INTRODUCTION

The rapid evolution of Artificial Intelligence (AI) has fundamentally transformed industries worldwide, leading to a surge in demand for AI professionals. Organizations are racing to integrate AI technologies into their operations, and job seekers are increasingly drawn to the promise of high-paying, future-proof careers in this field. However, this excitement is often clouded by misinformation, unrealistic expectations, and a lack of clarity around actual market trends. While many are aware of the AI boom, few have access to data-driven insights into how job roles, geographic location, and experience levels affect salary and career opportunities in AI.

In this context, individuals face several challenges: What are the realistic salary expectations for AI roles? How does experience level influence compensation? Which countries or regions offer the best opportunities? What roles are in highest demand? These questions are crucial for students, professionals, and career changers looking to enter or grow in the AI sector.

This project, *Global AI Job Market Salary Trends and Career Planning*, addresses these challenges by leveraging real-world job market data. The system uses machine learning to predict salaries based on key features such as job title, experience level, employment type, and geographic location. It also classifies salary levels into categories (High, Medium, Low) to make insights more actionable and user-friendly. Alongside prediction and classification models, the project includes market trend analysis, career guidance based on salary benchmarks, and geographic insights that visualize global AI employment landscapes.

The primary aim is to create a realistic and informative platform that demystifies the AI job market. Unlike the common narrative that all AI jobs are excessively high-paying or guaranteed, this tool uncovers the diversity in compensation and demand across different roles and regions. By doing so, it empowers users to make strategic decisions about their careers, rather than blindly following trends.

Ultimately, this project serves as a bridge between the booming world of AI and the individuals striving to find their place within it. Whether someone is a student exploring AI careers, a professional seeking a change, or a policymaker evaluating workforce trends, this system offers valuable, data-backed insights for informed decision-making.

2. LITERATURE SURVEY

2.1 Bankins et al. (2024) — Navigating Career Stages in the Age of Artificial Intelligence

Authors: Sarah Bankins, Stefan Jooss, Simon Lloyd D. Restubog, Mauricio Marrone, Anna Carmella Ocampo, Mindy Shoss

Source: *Journal of Vocational Behavior*, Vol. 153 (Open Access, CC BY-NC)

DOI: 10.1016/j.jvb.2024.104011

Highlights:

- Systematic review of 104 empirical studies on AI's impact on career trajectories
- Describes AI as an active agent shaping career exploration, guidance and management
- Employs a sustainable career lens, calling for AI systems that promote sustainable and equitable career development
- Addresses career stages theory to identify barriers, enablers, and impacts on career competencies.

2.2 Georgieff & Hyee (2021) — Artificial Intelligence and Employment: New Cross-Country Evidence

Authors: Alexandre Georgieff & Raphaela Hyee

Source: OECD Social, Employment and Migration Working Papers No. 265

DOI: 10.1787/c2c1d276-en

Highlights:

- Applies the Felten–Raj–Seamans AI Occupational Impact measure to 23 OECD countries
- Finds **no overall relationship** between AI exposure and employment growth (2012–2019)
- Shows that high computer-use occupations experienced **positive employment growth** with greater AI exposure
- In contrast, low computer-use occupations saw a **decline in average working hours** when exposed to AI
- Suggests **digital skills** as a buffer: tech-savvy workers shift to higher-value tasks and benefit from AI, while digitally less-skilled are at risk.

3. TOOLS AND TECHNOLOGIES USED:

This project leverages a mix of machine learning and web development tools to build a smart, intelligent, AI-career dashboard.

Category	Tools / Technologies	Purpose
Languages	Python	Primary programming language for ML, data processing, and Streamlit
Data Processing	pandas, numpy	Data wrangling, cleaning, and transformation
Machine Learning	scikit-learn	Used for modeling (classification, regression, clustering), pipelines, preprocessing
Model Persistence	joblib	Saving and loading ML models and encoders efficiently
Visualization	plotly.express, streamlit.components.v1	Interactive plots, maps, cluster cards
Web UI / Dashboard	Streamlit	UI layer for visualizing insights, interactive user input, and exporting results
Association Rules Mining	mlxtend (Apriori, association_rules)	Discovering patterns and rules from job data
Clustering Algorithms	KMeans, DBSCAN, AgglomerativeClustering	Unsupervised learning to group jobs based on salary, remote work, and experience
Dimensionality Reduction	PCA from sklearn.decomposition	For 2D visualization of high-dimensional cluster data
Encoding	LabelEncoder, OneHotEncoder, TransactionEncoder	Categorical data encoding for ML and association rule mining

4. SYSTEM DESIGN

4.1 SYSTEM ARCHITECTURE

The AI Career Scope system follows a modular architecture consisting of four main layers:

a. Frontend (User Interface Layer)

- Built using Streamlit, this layer allows users to interact with the system via a simple web interface.
- Users can input job-related details like position, location, experience, and skills.
- The frontend also displays the results: predictions, visualisations, salary graphs, and clustering outcomes.

b. Backend & Application Logic Layer

- This layer handles the business logic and connects the UI with the machine learning models.
- It processes the user input, formats it appropriately, and passes it to the relevant ML models.
- It then collects the predictions and computed insights and returns them to the UI for display.

c. Machine Learning Layer

This layer includes several pre-trained models saved as .pkl files using Joblib:

- Classification model (e.g., Gradient Boosting) predicts job categories.
- A regression model (e.g., Random Forest) estimates salary.
- Clustering model (e.g., KMeans, PCA) groups similar job profiles.
- Association Rule Engine (Apriori algorithm) finds patterns in job attributes.

These models are built and trained using scikit-learn, MLxtend, and NumPy.

d. Data Processing & Storage Layer

- The data (usually CSV files) is loaded and preprocessed using Pandas and scikit-learn's OneHotEncoder.
- This layer cleans, transforms, and encodes input features for model compatibility.
- It also stores model files (.pkl) and datasets needed for predictions and training.

4.2 SYSTEM WORKFLOW

Here is the logical flow of how the system works from start to finish:

a. User Input Stage

- The user accesses the Streamlit web app.

- They enter or upload job-related data such as job title, work mode, company size, education level, and experience.

b. Data Preprocessing Stage

- The input data is cleaned and transformed using Pandas.
- Categorical data is encoded, numerical values are scaled if needed, and missing values are handled.

c. Model Prediction Stage

The processed data is passed into the appropriate pre-loaded machine learning models:

- The Classification Model predicts the most suitable job role or career path.
- The Regression Model estimates the expected salary based on the input.
- The Clustering Model (like KMeans + PCA) groups the job profile into a cluster for comparison with similar roles.
- The Association Rule Model finds patterns like which benefits are common for remote jobs or what skills are most requested.

d. Visualisation and Insights Stage

- Results are returned to the user interface using Plotly charts, PCA scatter plots, and colour-coded cluster cards.
- Users can view salary trends, job clusters, and get career recommendations.
- Additional insights, such as top skills, common job types, and potential job growth, are displayed.

e. User Decision Making

The user uses the predictions and visualisations to understand:Career opportunities, Salary expectations, Skills to develop and Similar job profiles and industry trends.

5. WORK PLAN

The work plan outlines the sequential steps and milestones followed to complete the Global AI Market and Trend Analysis project efficiently and systematically. It ensures that all phases of development are addressed, from research and data analysis to implementation and deployment.

Phase 1: Requirement Analysis

- Defined the objectives: Analyze global AI job market trends using ML techniques and provide salary prediction, clustering, and association rule insights.
- Chose tools and technologies:
    - **ML backend:** pandas, scikit-learn, mlxtend
    - **Web framework:** Flask
    - **Frontend:** HTML, CSS, JavaScript
    - **Visualization:** Plotly, Streamlit

Phase 2: Data Preparation

- Collected and explored the csv dataset
- Cleaned the dataset by handling missing values and irrelevant columns.
- Scaled numerical features where needed for clustering.
- Split dataset into training and testing sets for supervised models.

Phase 3: Model Development

- **Regression:** Built a Random Forest Regressor to predict salaries based on features like job title, experience level, and remote ratio.
- **Clustering:** Applied K-Means to group job roles based on salary, remote work ratio, and benefits.
- **Association Rule Mining:** Used Apriori algorithm to discover patterns and frequent job trends.
- Evaluated models using appropriate metrics like R^2, MSE, and silhouette score.
- Serialized models using joblib and pickle for deployment.

Phase 4: Backend Integration

- Built a Flask-based backend to manage:
    - API routes for predictions and visualizations.
    - Integration of serialized models with endpoints.
    - Communication with frontend forms and user inputs.

- Ensured smooth data flow between HTML pages and ML model outputs.
- Added exception handling for model input and output errors.

Phase 5: Frontend Development

- Designed responsive HTML/CSS pages for:
    - Home Page
    - Salary Prediction
    - AI Career Planning
    - Market Trends
    - Geographic Studies
- Used JavaScript to handle interactivity and dynamic input rendering.
- Integrated charts and graphs using Plotly and embedded iframes.

Phase 6: Testing & Debugging

- Conducted thorough testing of:
    - Frontend navigation and responsiveness
    - Backend APIs and model predictions
    - Clustering visuals and rule mining outputs
- Resolved issues related to CORS, file paths, and data formatting.
- Validated model performance using test data.

Phase 7: Documentation and Reporting

- Compiled introduction, literature survey, system design, toolset, and implementation.
- Documented the entire workflow in this project report.
- Created a deployment-ready version for demo purposes using Streamlit and Flask.

6. CHALLENGES AND LIMITATIONS

a. Data Imbalance:

The dataset had a significant class imbalance (e.g., very few "Low" salary records), which affected the classification model's ability to perform well across all classes.

b. Incomplete or Inconsistent Data:

Many job listings had missing or inconsistent values (e.g., job titles, locations, experience levels), requiring extensive preprocessing and cleaning.

c. Dynamic Job Market:

Salaries and job trends change rapidly, especially in AI. Static datasets cannot fully capture real-time market shifts, limiting the model's long-term accuracy.

d. Geographic and Currency Variations:

Salary data from different countries with different currencies and cost of living factors may introduce inconsistency in predictions.

e. Generalization Limitations:

The models may not generalize well to unseen or highly niche roles due to limited representation in the training data.

f. Model Interpretability:

Advanced models like XGBoost and CatBoost offer high accuracy but are less interpretable, making it harder for users to understand how predictions are made.

g. Bias in Data:

The dataset may contain biases (e.g., gender, region, job title naming conventions), which could reflect in the model's predictions unknowingly.

7. FUTURE ENHANCEMENTS

a. **Real-Time Market Updates**
   Integrate APIs or web scraping to fetch real-time AI job postings, salaries, and demand trends from platforms like LinkedIn or Indeed.
b. **Personalized Career Roadmap**
   Recommend career paths and upskilling resources (like courses or certifications) based on the user's background, interests, and market demand.
c. **Interactive Geographic Visualization**
   Add a map-based interface showing region-wise salary heatmaps, role availability, and growth opportunities.
d. **Resume Analysis & Job Fit Score**
   Allow users to upload resumes and get AI-powered feedback on job-role alignment and suggestions to improve employability.
e. **Multi-language Support**
   Add support for regional languages to make the tool accessible to a broader audience, especially fresh graduates across India.
f. **AI Chatbot for Career Advice**
   Build a conversational assistant to answer questions about roles, salaries, or required skills, offering a more guided user experience.

8. CONCLUSION

This project aimed to bridge the gap between the growing interest in AI careers and the actual trends in the job market by leveraging data-driven techniques. Through the use of regression and classification models, we successfully predicted salary ranges and categorized job roles based on various attributes such as experience, location, and job title. The insights generated offer valuable guidance for students, professionals, and career switchers who are exploring opportunities in the AI field.

Beyond technical prediction, the project also emphasized career planning by highlighting high-demand roles and realistic compensation expectations across different geographies. This helps dispel common myths driven by AI hype and brings transparency to the job market. Our analysis demonstrates that while AI offers great potential, the benefits are influenced by multiple real-world factors like skill level, job location, and experience.

In conclusion, this system can serve as a helpful reference tool for individuals making career decisions, educational institutions guiding students, and organizations studying talent trends. With future enhancements like real-time data integration and personalized role matching, this platform can evolve into a more robust career intelligence system, empowering users to navigate the AI job landscape more strategically

9. REFERENCES

[1] Kaggle Datasets – AI Job Market & Salary Data 2025
https://www.kaggle.com/datasets

[2] Glassdoor – Tech Salaries and Job Insights
https://www.glassdoor.com

[3] LinkedIn Talent Insights – AI & Data Science Hiring Trends
https://business.linkedin.com/talent-solutions/talent-insights

[4] Indeed Salary Guide – Machine Learning and AI Roles
https://www.indeed.com/salaries

[5] Scikit-learn Documentation – Machine Learning Algorithms
https://scikit-learn.org/stable/documentation.html

[6] XGBoost Documentation – Gradient Boosting Techniques
https://xgboost.readthedocs.io

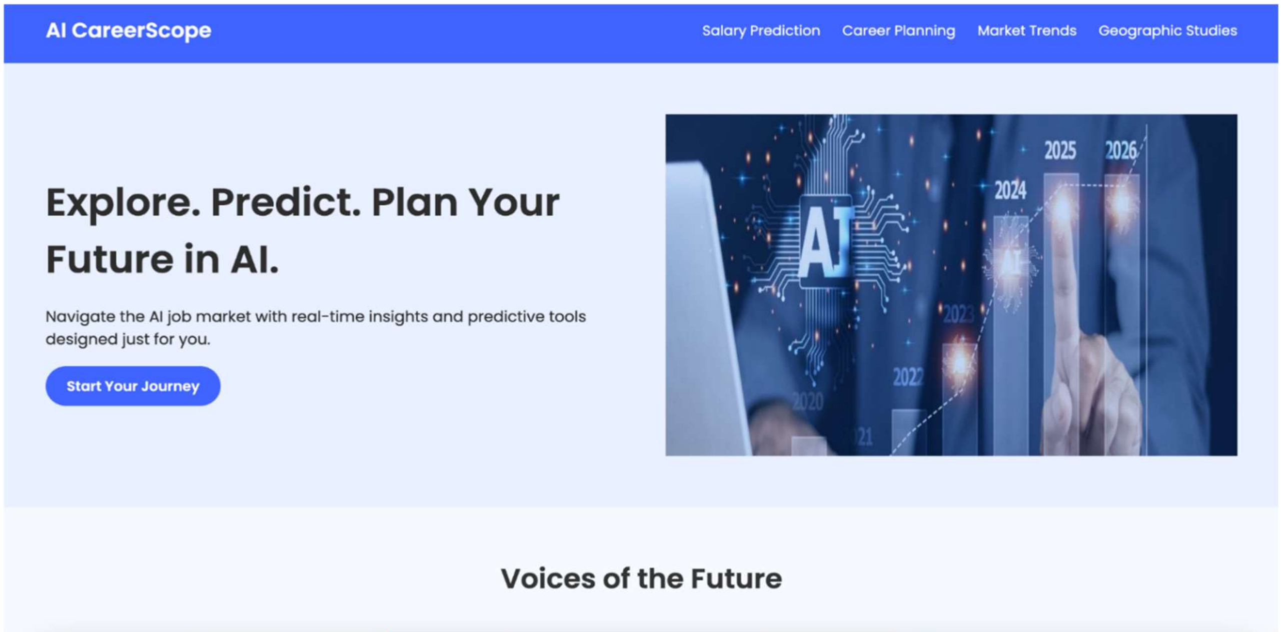
[7] LightGBM Documentation – Efficient Tree-Based Learning
https://lightgbm.readthedocs.io

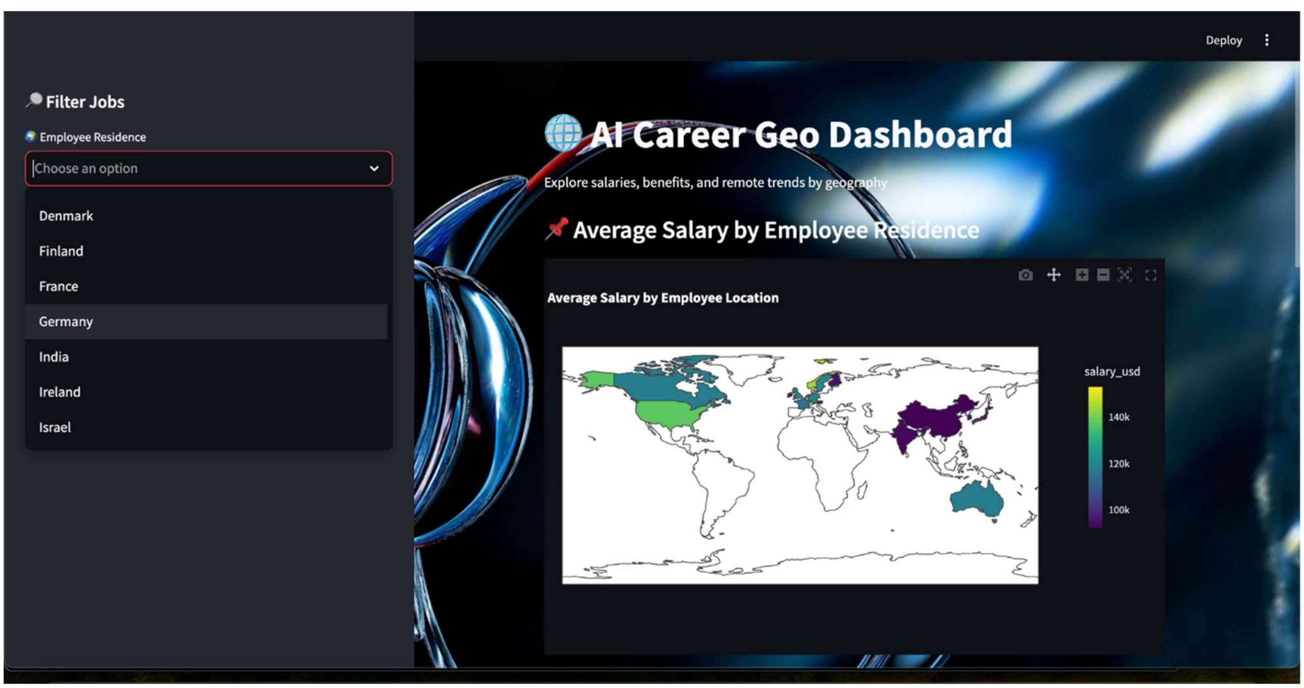
[8] CatBoost Documentation – Handling Categorical Features in ML
https://catboost.ai/docs

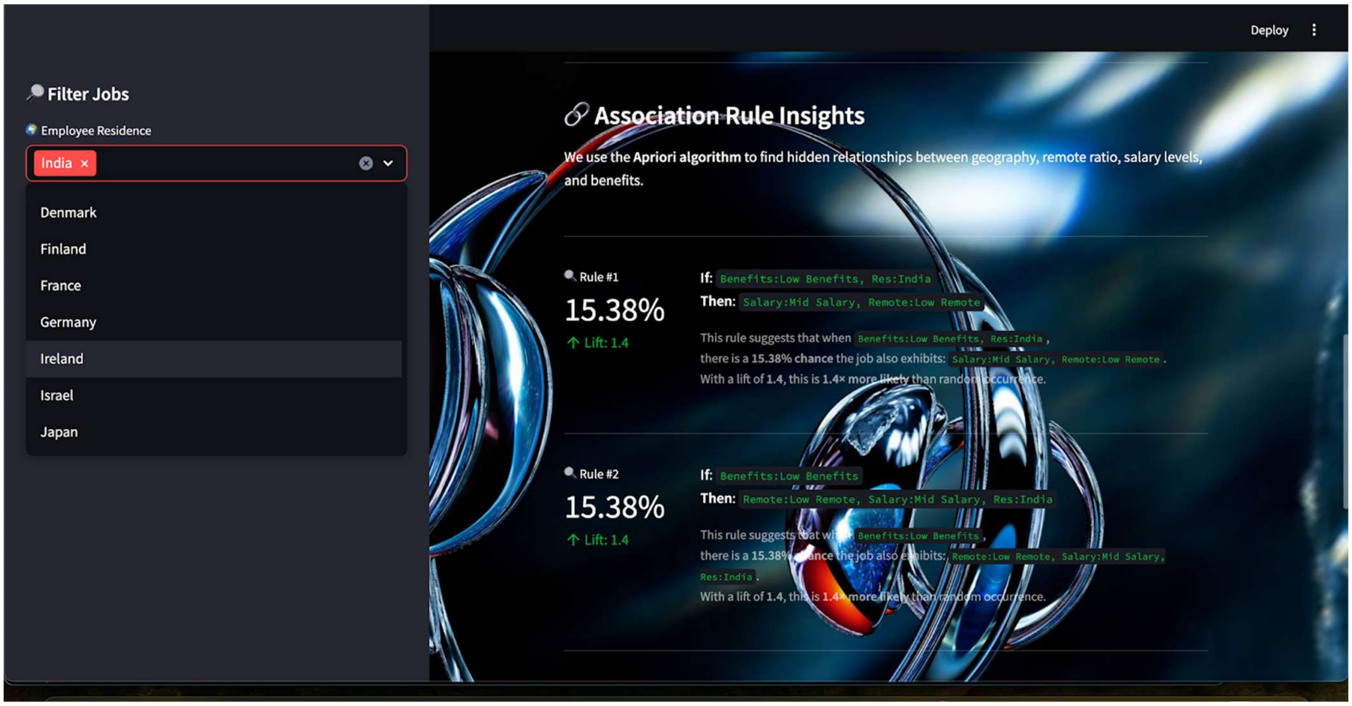
[9] Statista – AI Job Growth Projections Worldwide
https://www.statista.com

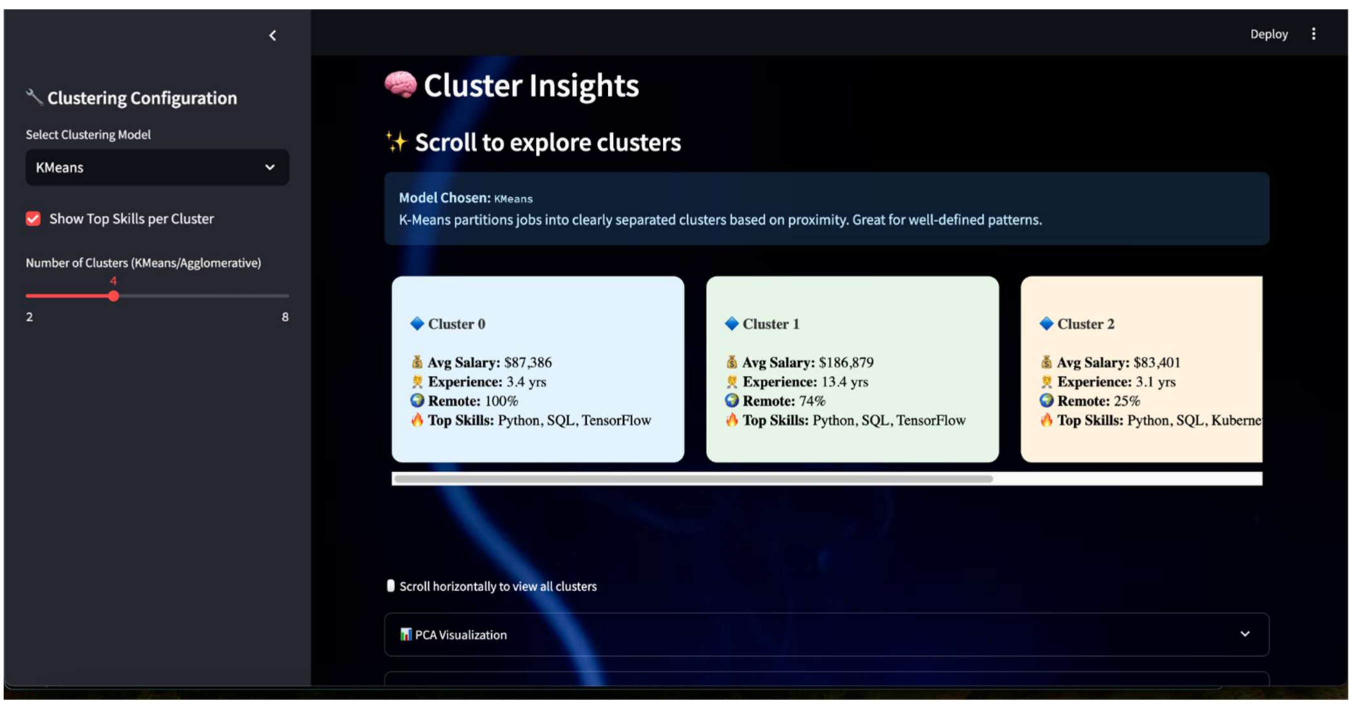
[10] World Economic Forum– The Future of Jobs Report
https://www.weforum.org/reports/the-future-of-jobs-report/

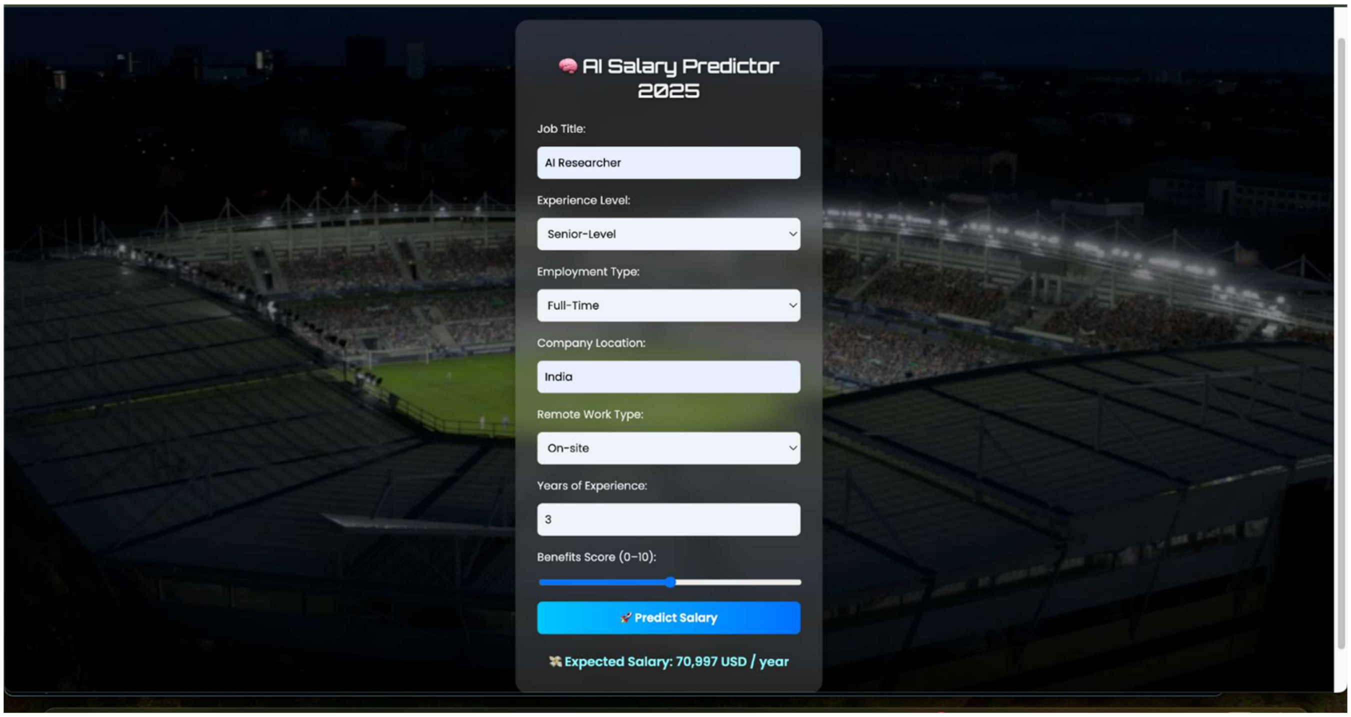
10. SNAPSHOTS

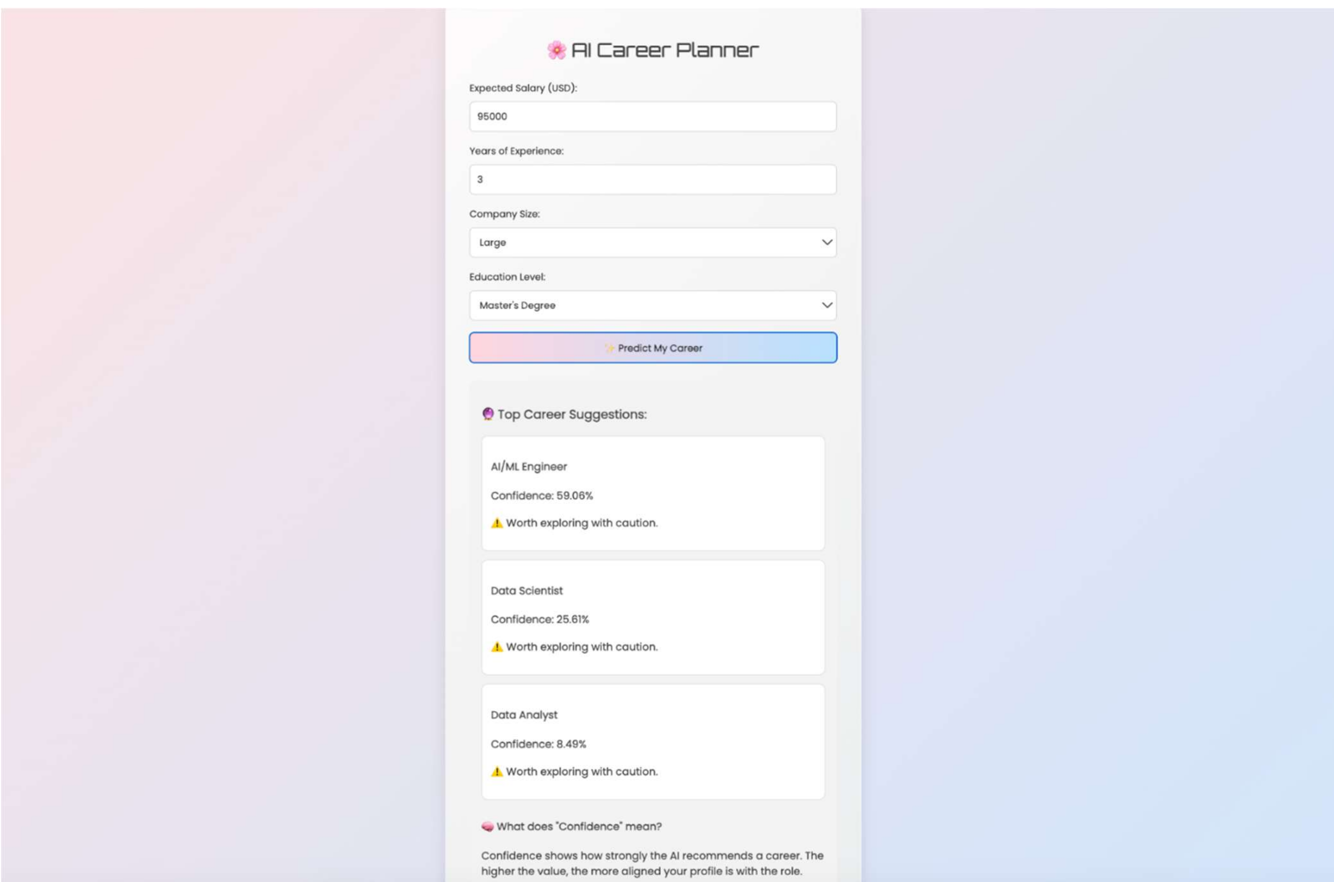





Deploy
Filter Jobs
Employee Residence
India ×
Denmark
Finland
France
Germany
Ireland
Israel
Japan
Association Rule Insights
We use the Apriori algorithm to find hidden relationships between geography, remote ratio, salary levels, and benefits.
Rule #1
15.38%
Lift: 1.4
If: Benefits:Low Benefits, Res:India
Then: Salary:Mid Salary, Remote:Low Remote
Rule #2
15.38%
Lift: 1.4
If: Benefits:Low Benefits
Then: Remote:Low Remote, Salary:Mid Salary, Res:India


Deploy
Clustering Configuration
Select Clustering Model
KMeans
Show Top Skills per Cluster
Number of Clusters (KMeans/Agglomerative)
4
2
8
Cluster Insights
Scroll to explore clusters
Model Chosen: KMeans
K-Means partitions jobs into clearly separated clusters based on proximity. Great for well-defined patterns.
Cluster 0
Avg Salary: $87,386
Experience: 3.4 yrs
Remote: 100%
Top Skills: Python, SQL, TensorFlow
Cluster 1
Avg Salary: $186,879
Experience: 13.4 yrs
Remote: 74%
Top Skills: Python, SQL, TensorFlow
Cluster 2
Avg Salary: $83,401
Experience: 3.1 yrs
Remote: 25%
Scroll horizontally to view all clusters
PCA Visualization


AI Salary Predictor 2025
Job Title:
AI Researcher
Experience Level:
Senior-Level
Employment Type:
Full-Time
Company Location:
India
Remote Work Type:
On-site
Years of Experience:
3
Benefits Score (0-10):
Predict Salary
Expected Salary: 70,997 USD / year


AI Career Planner
Expected Salary (USD):
95000
Years of Experience:
3
Company Size:
Large
Education Level:
Master's Degree
Predict My Career
Top Career Suggestions:
AI/ML Engineer
Confidence: 59.06%
Worth exploring with caution.
Data Scientist
Confidence: 25.61%
Worth exploring with caution.
Data Analyst
Confidence: 8.49%
Worth exploring with caution.
What does "Confidence" mean?
Confidence shows how strongly the AI recommends a career. The higher the value, the more aligned your profile is with the role.