WAGEWIZARD

An AI-Powered Guide to Your First Paycheck

Presented by:
Ruchira Lokhande, Garima Mathur, Varun Rao, Amit Gangane, Priyanka Iyer, Pooja Kumari

Table of Contents

01 Business Questions

02 Proposed Solution

03 ML Pipeline

04 The Dataset

05 EDA

06 Model Validation Metrics

07 Results

08 Machine Learning Canvas

09 Conclusion & Next Steps

What Are We Trying to Predict?

Unclear Salary Expectations

Job Market Anxiety

Career Planning Confusion

No Personalized Insights

To address the challenges graduates and job seekers face, our project seeks to answer the following 3 questions:

How can we accurately predict the salary of graduates and job seekers using their academic and professional profile?

Which factors have the greatest influence on a graduate's or job seeker's learning potential?

Can we offer personalized, data-driven insights to guide students in improving their profiles?

From Guesswork to Guidance: Our AI-Driven Salary Predictor

WageWizard is a machine learning-based predictive model developed to estimate the salary of graduates and job seekers. Using input features like education level, degree type, academic performance, and prior experience, the model aims to provide accurate salary range predictions.

PREDICTION MODEL

End-to-end machine learning pipeline to estimate salaries for graduates and job seekers.

PERSONALIZED INSIGHTS

Provides personalized salary estimates based on academic and experience profiles

SEAMLESS ML WORKFLOW

Includes preprocessing, model training, evaluation, and interpretable outputs

From Data to Decision: The ML Journey

Data Preprocessing

Cleaned and prepared the dataset by handling missing values, encoding categorical features, and scaling numerical variables to ensure consistency for model training.

Exploratory Data Analysis

Explored feature distributions, patterns, and correlations to identify key drivers of salary and understand the impact of academic and experiential attributes.

Model Building & Evaluation

Trained regression models including Linear Regression, Random Forest, and XGBoost. Evaluated them using R^2, MAE, and RMSE to select the best model.

Final Model Deployment

Finalized and saved the best-performing model, preparing it for future integration into an interactive application using Streamlit.

The Dataset: Real-World Profiles for Realistic Predictions

Sourced from Kaggle's curated Salary Dataset, this dataset provides real-world job and education profiles to help us model salary expectations for job seekers entering the workforce or changing jobs.

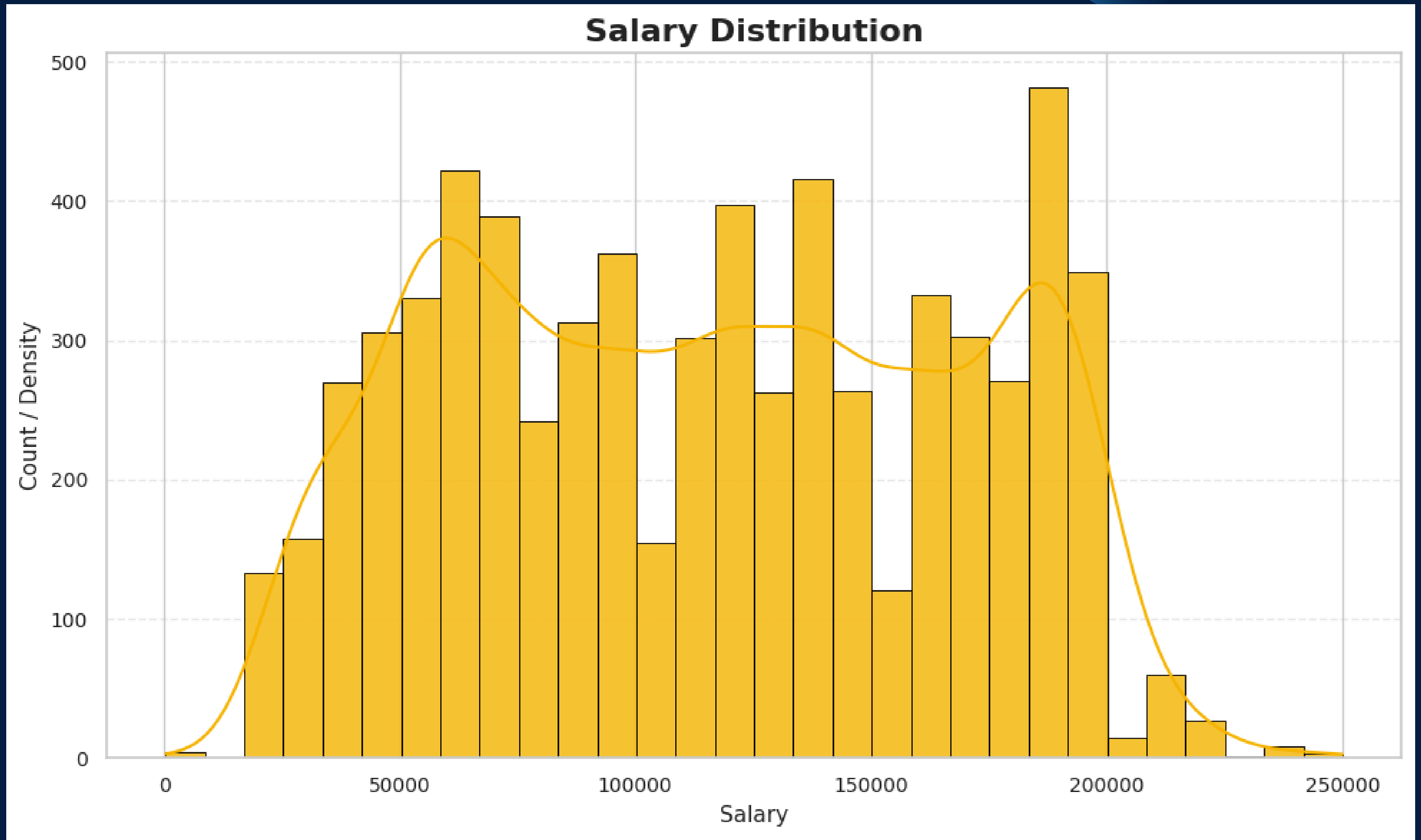
We selected it for its:

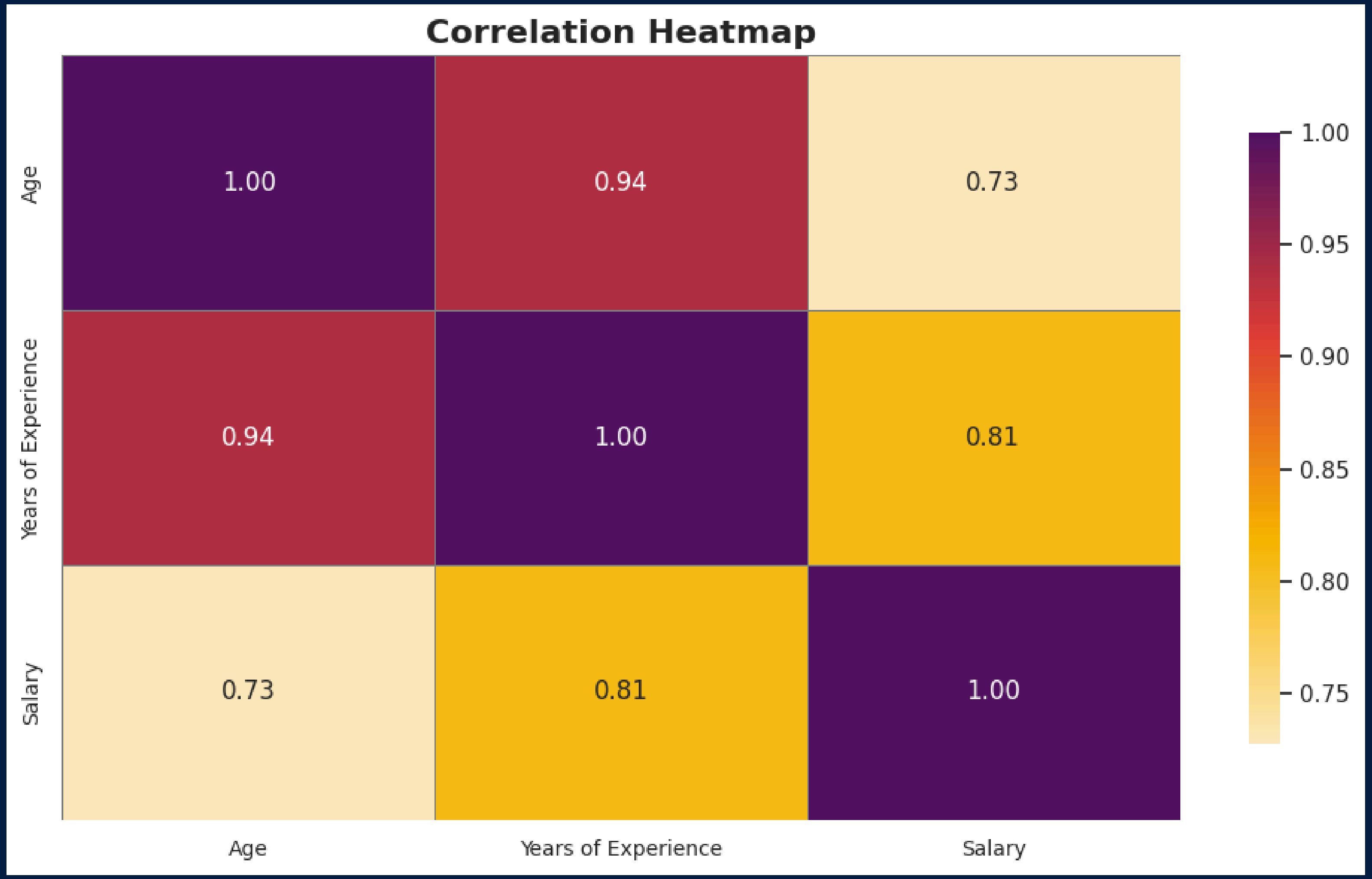
- Breadth (6,704 complete records)
- Clean structure suitable for modeling
- Relevant features aligning with our project goals

Though pre-collected, this dataset mirrors the realities of graduate and professional hiring, making it ideal for training models grounded in real-world outcomes.

FEATURES	DESCRIPTION
Age	Age of the graduate at time of job placement
Education Level	From High-School to PhD
Experience	Prior internships, freelance, or part-time
Job Role	Entry-level title obtained
Salary	Annual compensation in USD

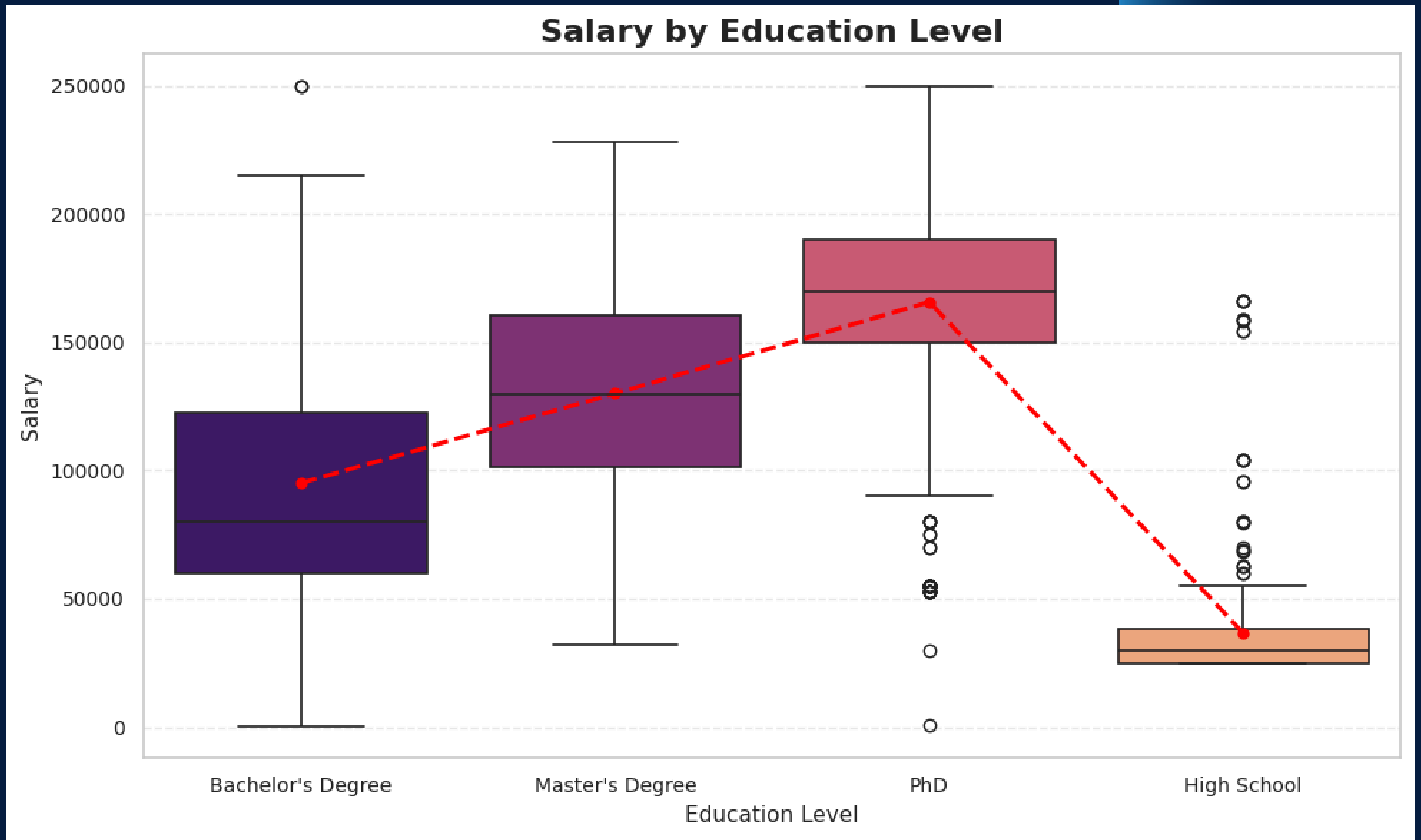
Cracking the Code:
What Influences a Salary?

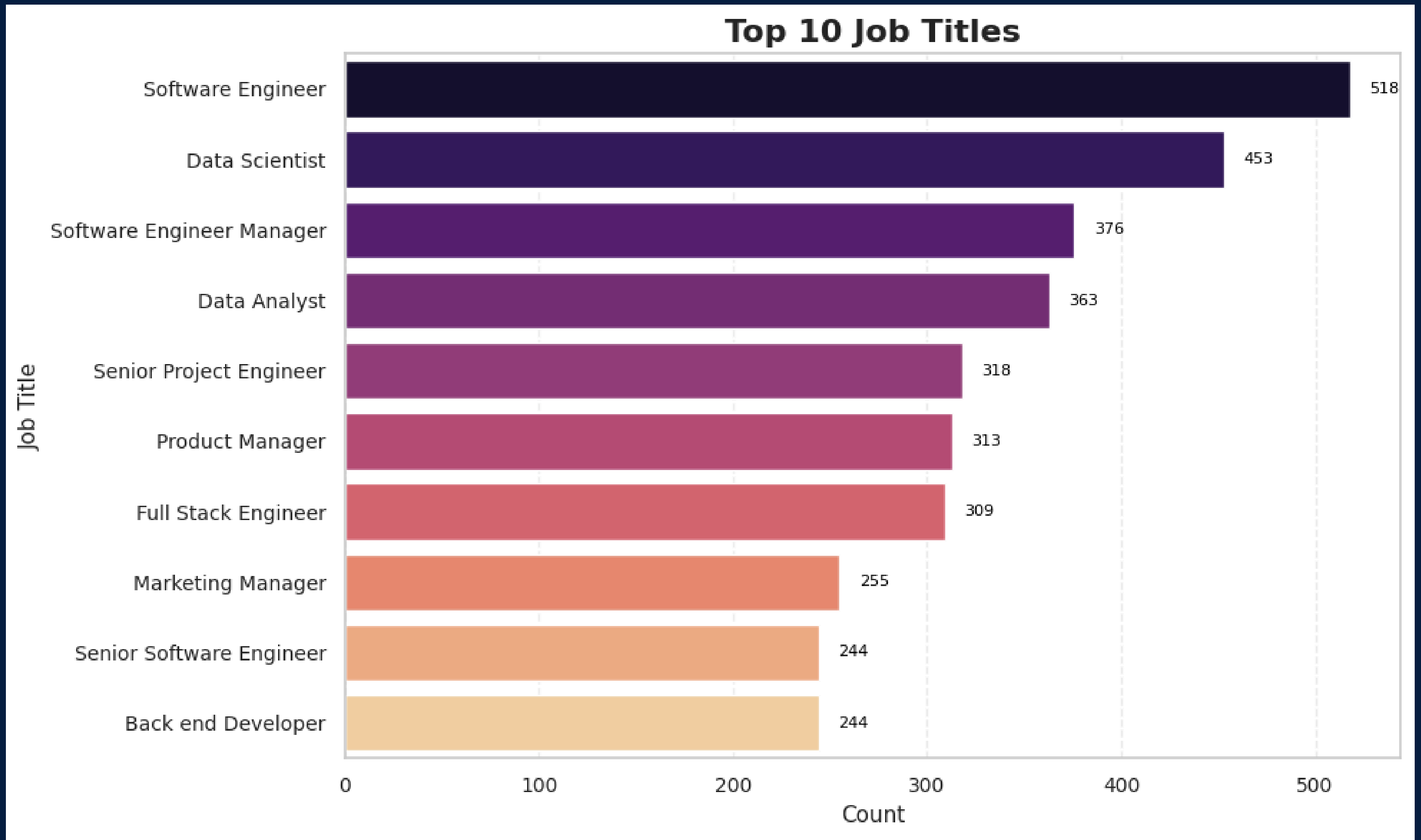




We started by understanding how salaries are distributed and what variables correlate most strongly. Experience, unsurprisingly, emerged as the strongest predictor—more than even age

Who Earns What:
Role & Education Insights





We then examined how education and job roles influence salaries. Master's and PhD holders tend to earn more on average, and roles like Software Engineer and Data Scientist dominate the job market.

Can We Trust the Prediction?

To predict graduate salaries, we tested three algorithms using structured data.

Linear Regression Performance:

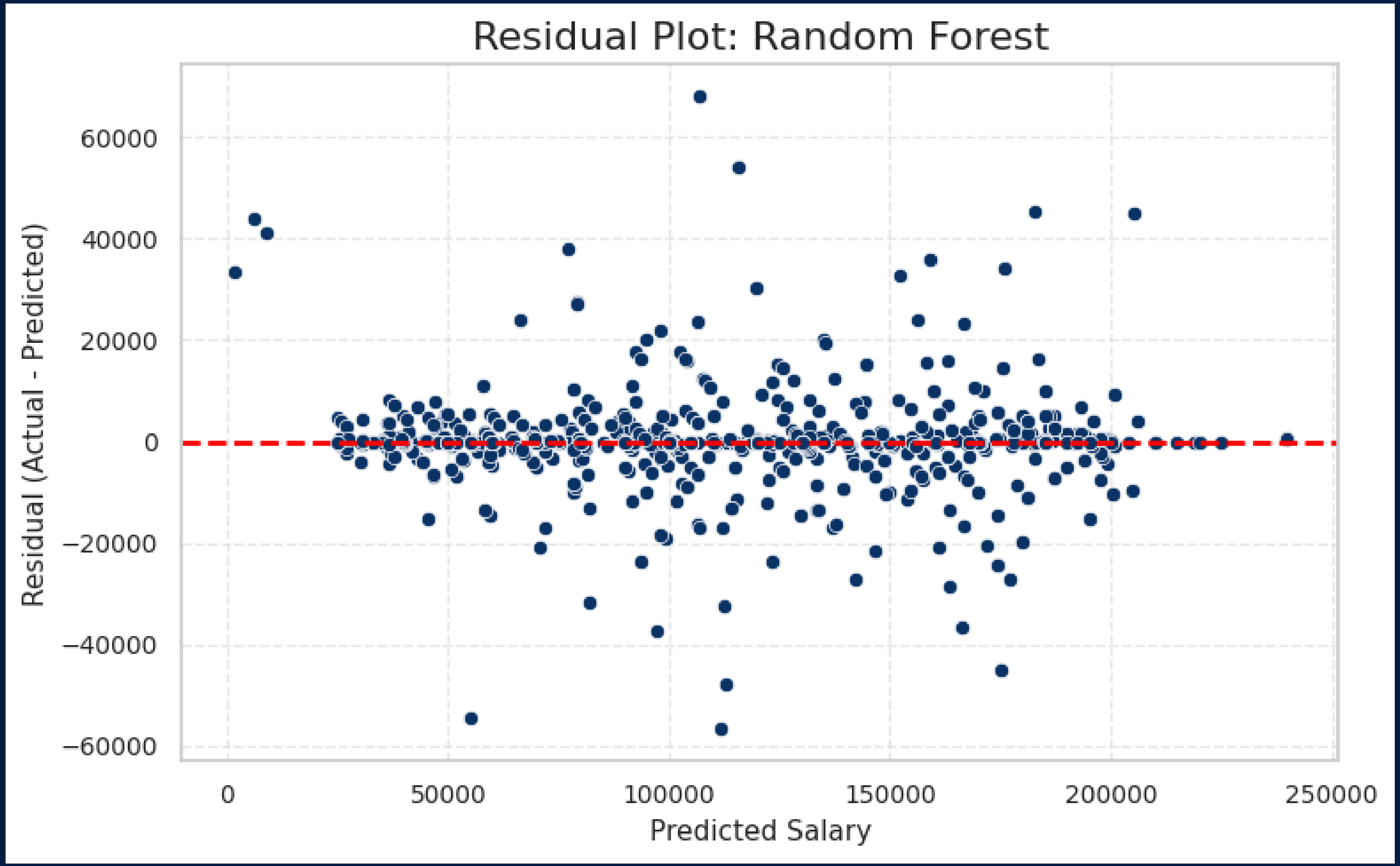
- MAE: $15,073.16
- RMSE: $24,332.75
- R^2: 0.7924

Random Forest Performance:

- MAE: $2,880.43
- RMSE: $7,333.68
- R^2: 0.9811

XGBoost Performance:

- MAE: $5,066.92
- RMSE: $8,456.83
- R^2: 0.9749



Random Forest produced the best results, with the lowest MAE and RMSE, and the highest R^2.
This aligns with expectations for structured data and made it our final model choice.

How Well Did WageWizard Predict?

Using a few key inputs—education level, job title, age, and work experience—WageWizard delivers tailored salary estimates within seconds.
These examples highlight how small differences in profile can lead to significant shifts in earning potential.

```
Extracted Resume Info:
```

Age	Gender	Education Level	Job Title	Years of Experience
24	female	master's	Business Analyst	2

```
==================================================
PREDICTED STARTING SALARY:  $65,035.52
==================================================
```

A **Master's-level Business Analyst** with 2 years of experience is estimated to earn **$65K**, aligning with industry benchmarks for early-career roles focused on data-driven decision-making, reporting, and team support.

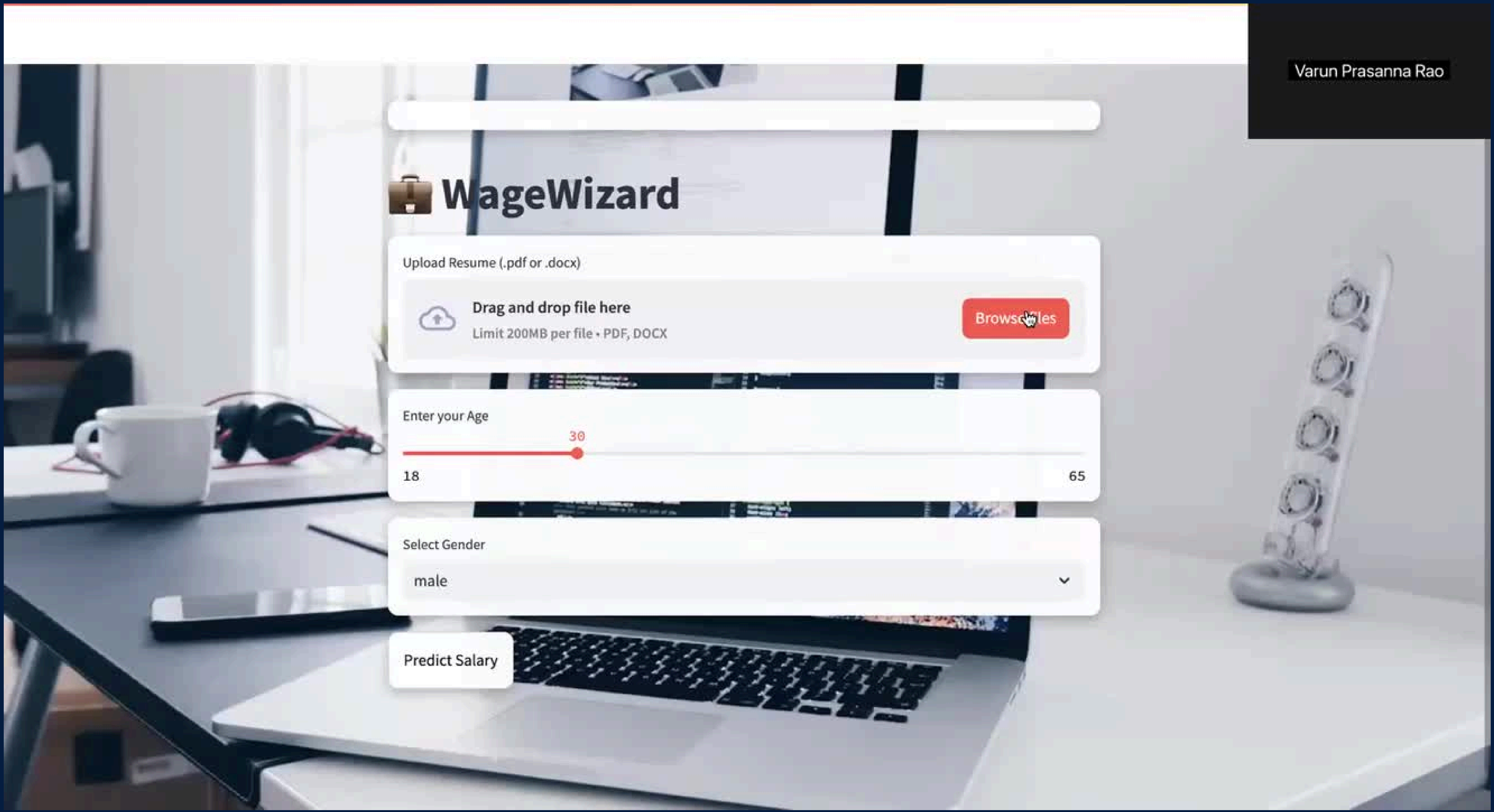
A **PhD-level Product Manager** with 7 years of experience is projected to earn **$145K**. This reflects the market's premium on professionals who combine advanced academic credentials with years of cross-functional leadership and product strategy expertise.

```
Extracted Resume Info:
```

Age	Gender	Education Level	Job Title	Years of Experience
30	male	phd	Product Manager	7

```
==================================================
PREDICTED STARTING SALARY:  $145,449.50
==================================================
```

WageWizard in Action!



From Data to Decisions: Our ML Canvas

PREDICTION TASK

- **Type of task**: Supervised Regression
- **Entity**: Individual student or job-seeker profile
- **Predicted outcome**: Annual starting salary in USD
- **When observed**: After the individual secures a job and receives a salary offer

DECISIONS

Actionable output: Predicted salary is used to:

- Set realistic expectations for students
- Help career counselors offer tailored advice
- Guide students toward skill-building decisions (e.g., gaining internships or pursuing further education)

How: Via an interactive web app where users enter profile details (age, education, job title, experience) and instantly receive a salary prediction

Parameters: Real-time input, instant response, and suggested profile improvements based on salary gaps

VALUE PROPOSITION

End beneficiary: Fresh graduates, jobseekers, university career services, training institutes

Pain points addressed:

- Unclear salary benchmarks
- Mismatched expectations in career planning
- Lack of evidence-based career guidance

Integration & UI:

- Streamlit-based web interface for individual use
- Future integration with university career portals or job prep platforms
- Potential chatbot or API for guided Q&A salary predictions

DATA COLLECTION

Initial sourcing:

- Kaggle's publicly available structured Salary Dataset
- Real-world job and education data, manually curated and cleaned

Update strategy:

- Periodic scraping or API integration with job boards (e.g., LinkedIn, Indeed) in future iterations
- Consider cost-effective crowdsourced validation or surveys for fresh graduate outcomes

DATA SOURCES

External sources:

- Kaggle Salary Dataset (structured CSV)
- Future potential: APIs from LinkedIn, Glassdoor, PayScale, or government wage data

Internal sources (future):

- Student databases from academic institutions
- Career services intake forms

IMPACT SIMULATION

Cost/gain of correct predictions: Accurate predictions improve student trust, better job targeting, and resource alignment

Cost of incorrect predictions:

- Underestimation could demotivate students
- Overestimation might lead to unrealistic goals or dissatisfaction

Pre-deployment simulation:

- Validated model using holdout test sets (20%) and cross-validation

Deployment criteria:

- Minimum $R^2 > 0.95$
- MAE under $5,000

Fairness constraints:

- Avoid bias based on gender, race, or age (planned in future iterations)
- Audit feature contributions using SHAP to detect unfair influence

MAKING PREDICTIONS

Mode: On-demand (real-time) via Streamlit UI

Frequency: As often as user submits input

Latency: Few milliseconds (lightweight Random Forest inference)

Resources: Local CPU for Streamlit app, scalable to cloud GPU/VM instance if hosted

MONITORING

KPIs (Model-centric):

- R^2 score > 0.95
- MAE < $3,000
- Low variance across demographic groups

KPIs (Business-centric):

- % of users reporting improved career planning
- % of students following advice and reporting higher satisfaction
- Number of accurate predictions verified by actual salaries (future phase)

Review cadence:

- Monthly model performance reports
- Quarterly feedback loops from users and institutions

BUILDING MODELS

Number of models: One primary regression model (Random Forest Regressor)

Update frequency:

- Retrain quarterly or when new data is acquired
- Reassess feature relevance every 6 months

Time constraints: Short training time (~minutes), retrain offline

Resources: Local machine or cloud-based training using scikit-learn or XGBoost pipelines

FEATURES

Number of models: One primary regression model (Random Forest Regressor)

Update frequency:

- Retrain quarterly or when new data is acquired
- Reassess feature relevance every 6 months

Time constraints: Short training time (~minutes), retrain offline

Resources: Local machine or cloud-based training using scikit-learn or XGBoost pipelines

What's Next for WageWizard?

CONCLUSION

WageWizard delivers clear, personalized salary predictions using structured profile data.

The Random Forest model achieved an impressive R^2 of 0.98, demonstrating strong predictive performance on unseen data and confirming its reliability for real-world application.

Beyond predictions, WageWizard lays the groundwork for an intelligent career support tool—one that can evolve with new data, technologies, and student needs.

NEXT STEPS

Expand Inputs:
Add GPA, Certifications, and location data.

Continuous Model Retraining
Fine-tune the model regularly using new job market data from platforms like LinkedIn, Glassdoor, or Payscale to keep predictions relevant and aligned with real-world salary shifts.

Benchmark Against Real Offers:
Let users anonymously submit job offers to compare with predictions and fine-tune model accuracy using crowd-sourced real-world data.

THANK YOU

WageWizard can’t predict your boss... but it can predict your paycheck!

Presented by:
Ruchira Lokhande, Garima Mathur, Varun Rao, Amit Gangane, Priyanka Iyer, Pooja Kumari