

Finding job Prediction During Pathrise Program

K Nearest Neighbor

Delaram Doroudgarian
(5881909)

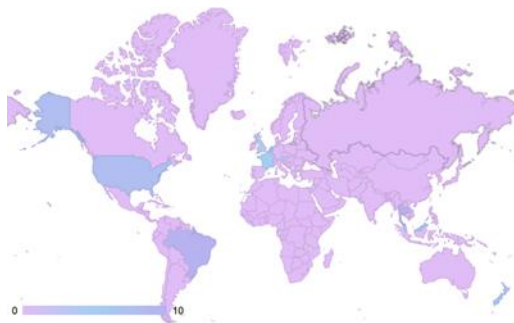
Seyede Aida
Atarzadeh Hosseini
(6936800)



2024/2025

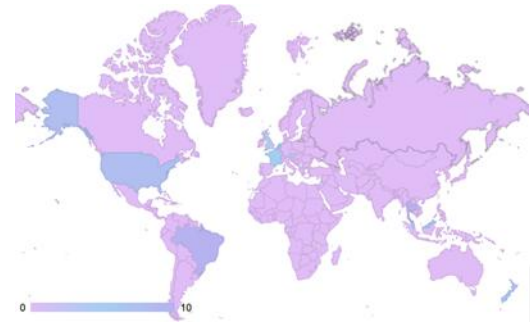
INTRODUCTI ON

In this analysis, we explore data from Pathrise, an online program aiding job seekers in the tech industry. The primary focus is to predict if a fellow will be placed at a company.



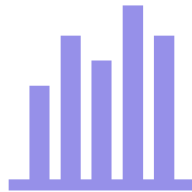
The main question that needs to be answered in this project is:

- Can we predict if candidate will be placed at Pathrise?



Dataset

Participant information of the Pathrise program



The dataset contains information including high level of education, length of job search, professional experience, and more for 2544 individuals, and the "placed" column represents the main outcome, indicating whether a person was hired (1) or not (0) during the program



| Type | Column | Row |
|--------|--------|------|
| All | 17 | 2544 |
| Object | 11 | 2544 |
| Int | 4 | 2544 |
| Float | 2 | 2544 |

Data Overview

| # | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q |
|----|----|----------------------|---------------|------------|-----------------------|--------|--------------------|----------------------------|----------------------|--------------------------------------|-------------------------|---------------------------|----------------------|------------------------|------------|------------|---|
| 1 | id | pathrise_status | primary_track | cohort_tag | program_duration_days | placed | employment_status | highest_level_of_education | length_of_job_search | biggest_challenge_in_search | professional_experience | work_authorization_status | number_of_interviews | number_of_applications | gender | race | |
| 2 | 0 | 1 Active | SWE | OCT19A | | 0 | Unemployed | Bachelor's Degree | 3-5 months | Hearing back on my applications | 3-4 years | Canada Citizen | 2 | | 900 Male | Non-Hisp | |
| 3 | 1 | 2 Active | PSO | JAN20A | | 0 | Unemployed | Some College, No Degree | 3-5 months | Getting past final round interviews | 1-2 years | Citizen | 6 | | 0 Male | Non-Hisp | |
| 4 | 2 | 3 Closed Lost | Design | AUG19B | | 0 | Employed Part-Time | Master's Degree | Less than one month | Figuring out which jobs to apply for | Less than one year | Citizen | 0 | | 0 Male | East Asia | |
| 5 | 3 | 4 Closed Lost | PSO | AUG19B | | 0 | Contractor | Bachelor's Degree | Less than one month | Getting past final round interviews | Less than one year | Citizen | 5 | | 25 Male | Decline to | |
| 6 | 4 | 5 Placed | SWE | AUG19A | 89 | 1 | Unemployed | Bachelor's Degree | 1-2 months | Hearing back on my applications | 1-2 years | F1 Visa/OPT | 10 | | 100 Male | East Asia | |
| 7 | 5 | 6 Closed Lost | SWE | AUG19A | | 0 | Employed Full-Time | Master's Degree | 1-2 months | Technical interviewing | 3-4 years | Green Card | 5 | | 100 Male | East Asia | |
| 8 | 6 | 7 Closed Lost | SWE | AUG19B | | 0 | Employed Full-Time | Master's Degree | Less than one month | Getting past phone screens | 3-4 years | Green Card | 0 | | 9 Male | Black, Afr | |
| 9 | 7 | 8 Withdrawn (Failed) | SWE | AUG19A | | 19 | Employed Part-Time | Bachelor's Degree | Less than one month | Getting past final round interviews | 1-2 years | Citizen | 4 | | 15 Female | Latino or | |
| 10 | 8 | 9 Active | SWE | AUG19B | | 0 | Student | Master's Degree | Less than one month | Technical interviewing | 1-2 years | F1 Visa/CPT | 1 | | 5 Male | East Asia | |
| 11 | 9 | 10 Withdrawn (Trial) | SWE | SEP19A | 13 | 0 | Employed Full-Time | Master's Degree | Less than one month | Getting past final round interviews | 3-4 years | Citizen | 0 | | 10 Male | Black, Afr | |
| 12 | 10 | 11 Closed Lost | PSO | AUG19B | | 0 | Unemployed | Master's Degree | 1-2 months | Hearing back on my applications | 1-2 years | Other | 0 | | 3 Male | Latino or | |
| 13 | 11 | 12 Withdrawn | Data | AUG19C | 158 | 0 | Unemployed | Master's Degree | 3-5 months | Lack of relevant experience | 5+ years | Citizen | 5 | | 50 Male | Decline to | |
| 14 | 12 | 13 Withdrawn (Trial) | Design | OCT19A | | 12 | Contractor | Bachelor's Degree | 6 months to a year | Getting past phone screens | 1-2 years | Green Card | 3 | | 10 Male | Middle Ea | |
| 15 | 13 | 14 Withdrawn | PSO | OCT19A | | 52 | Employed Part-Time | Bachelor's Degree | 1-2 months | Lack of relevant experience | Less than one year | Citizen | 4 | | 40 Male | Non-Hisp | |
| 16 | 14 | 15 Active | PSO | DEC19A | | 0 | Employed Full-Time | Bachelor's Degree | 1-2 months | Technical skills | 1-2 years | Citizen | 2 | | 35 Male | Non-Hisp | |
| 17 | 15 | 16 Active | PSO | JAN20A | | 0 | Employed Full-Time | Bachelor's Degree | 3-5 months | Getting past mid-stage interviews | Less than one year | Citizen | 1 | | 25 Male | South Asi | |
| 18 | 16 | 17 Active | PSO | FEB20A | | 0 | Student | Bachelor's Degree | Less than one month | Getting past phone screens | 3-4 years | Green Card | 1 | | 35 Male | East Asia | |
| 19 | 17 | 18 Active | SWE | JAN20A | | 0 | Student | Bachelor's Degree | 1-2 months | Lack of relevant experience | 1-2 years | Green Card | 5 | | 45 Female | Middle Ea | |
| 20 | 18 | 19 Active | SWE | JAN20A | | 0 | Unemployed | Bachelor's Degree | 3-5 months | Hearing back on my applications | Less than one year | Citizen | | | 15 Male | East Asia | |
| 21 | 19 | 20 Withdrawn (Trial) | Data | AUG19A | 13 | 0 | Employed Full-Time | Bachelor's Degree | 1-2 months | Getting past final round interviews | 1-2 years | Citizen | 0 | | 70 Male | Latino or | |
| 22 | 20 | 21 Active | Data | FEB20A | | 0 | Employed Part-Time | Master's Degree | 3-5 months | Technical interviewing | 3-4 years | Citizen | 3 | | 30 Male | East Asia | |
| 23 | 21 | 22 Withdrawn (Trial) | SWE | NOV19A | | 11 | Student | Master's Degree | 3-5 months | Hearing back on my applications | 3-4 years | F1 Visa/OPT | 1 | | 9 Male | Black, Afr | |
| 24 | 22 | 23 Withdrawn | SWE | AUG19A | | 93 | Contractor | Bachelor's Degree | 6 months to a year | Hearing back on my applications | 1-2 years | Citizen | 1 | | 10 Female | East Asia | |
| 25 | 23 | 24 Placed | PSO | NOV19B | | 193 | 1 Unemployed | Master's Degree | 6 months to a year | Hearing back on my applications | 1-2 years | Citizen | 5 | | 4 Male | Non-Hisp | |
| 26 | 24 | 25 Closed Lost | Design | NOV19B | | 0 | Contractor | | 1-2 months | Hearing back on my applications | 3-4 years | Green Card | 0 | | 20 | East Asia | |
| 27 | 25 | 26 Active | SWE | JAN20B | | 0 | Student | Bachelor's Degree | 1-2 months | Lack of relevant experience | Less than one year | F1 Visa/OPT | 0 | | 130 Male | East Asia | |
| 28 | 26 | 27 Active | SWE | FEB20B | | 0 | Student | Bachelor's Degree | 1-2 months | Figuring out which jobs to apply for | 1-2 years | Citizen | 2 | | 15 Male | Latino or | |
| 29 | 27 | 28 Withdrawn (Trial) | SWE | NOV19A | | 12 | Unemployed | Bachelor's Degree | 3-5 months | Technical interviewing | 5+ years | Citizen | 0 | | 20 Female | East Asia | |
| 30 | 28 | 29 Closed Lost | PSO | JAN20A | | 0 | Employed Full-Time | Bachelor's Degree | 6 months to a year | Hearing back on my applications | Less than one year | Citizen | | | 45 Male | Non-Hisp | |
| 31 | 29 | 30 Withdrawn (Trial) | Design | JAN20B | | 13 | Unemployed | Master's Degree | 1-2 months | Getting past phone screens | 1-2 years | F1 Visa/OPT | 0 | | 110 Female | Latino or | |
| 32 | 30 | 31 Placed | SWE | NOV19A | 73 | 1 | Student | Master's Degree | Less than one month | Figuring out which jobs to apply for | 1-2 years | F1 Visa/CPT | 0 | | 1 Male | East Asia | |
| 33 | 31 | 32 Active | SWE | JAN20B | | 0 | Student | Bachelor's Degree | 3-5 months | Getting past phone screens | Less than one year | Other | 5 | | 10 Male | East Asia | |
| 34 | 32 | 33 Withdrawn | SWE | NOV19A | 286 | 0 | Student | Bachelor's Degree | 1-2 months | Hearing back on my applications | 1-2 years | F1 Visa/OPT | 0 | | 15 Female | East Asia | |
| 35 | 33 | 34 Active | SWE | NOV19A | | 0 | Employed Part-Time | Master's Degree | Less than one month | Technical interviewing | Less than one year | F1 Visa/CPT | 1 | | 3 Male | Black, Afr | |
| 36 | 34 | 35 Withdrawn (Trial) | SWE | JAN20B | 5 | 0 | Student | Some College, No Degree | 6 months to a year | Figuring out which jobs to apply for | 5+ years | Citizen | 5 | | 30 Female | Two or M | |
| 37 | 35 | 36 Active | Design | FEB20A | | 0 | Student | Bachelor's Degree | 1-2 months | Hearing back on my applications | 1-2 years | F1 Visa/OPT | 2 | | 15 Female | East Asia | |
| 38 | 36 | 37 Withdrawn (Trial) | Data | NOV19A | 19 | 0 | Student | Master's Degree | Less than one month | Behavioral interviewing | 3-4 years | F1 Visa/OPT | 3 | | 9 Female | East Asia | |
| 39 | 37 | 38 Withdrawn (Trial) | Design | DEC19A | 7 | 0 | Contractor | Master's Degree | 1-2 months | Hearing back on my applications | 1-2 years | Citizen | 6 | | 80 Female | East Asia | |
| 40 | 38 | 39 Active | PSO | NOV19B | | 0 | Contractor | Bachelor's Degree | 1-2 months | Getting past final round interviews | 5+ years | | 4 | | 25 Male | East Asia | |
| 41 | 39 | 40 Placed | SWE | NOV19A | 83 | 1 | Employed Full-Time | Master's Degree | 3-5 months | Getting past final round interviews | 5+ years | F1 Visa/OPT | | | 15 Male | Native Am | |
| 42 | 40 | 41 Active | PSO | DEC19A | | 0 | Unemployed | Master's Degree | 1-2 months | Getting past mid-stage interviews | Less than one year | Citizen | 20 | | 8 Female | Black, Afr | |



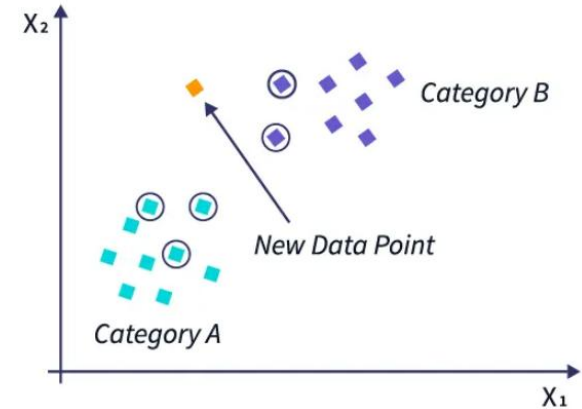
Provide Solution

For the question posed in the introduction section :

- Can we predict if candidate will be placed at Pathrise?

Leveraging Machine Learning

- Machine learning helps extract important information from data, leading to better decision-making.
- Machine learning models can learn from data and make decisions automatically and improve overtime.
- The KNN is a supervised learning classifier, which works well for simple classification tasks, making it ideal for small datasets.

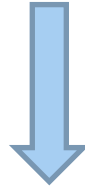




To make the dataset ready, it's needed to do
preprocessing

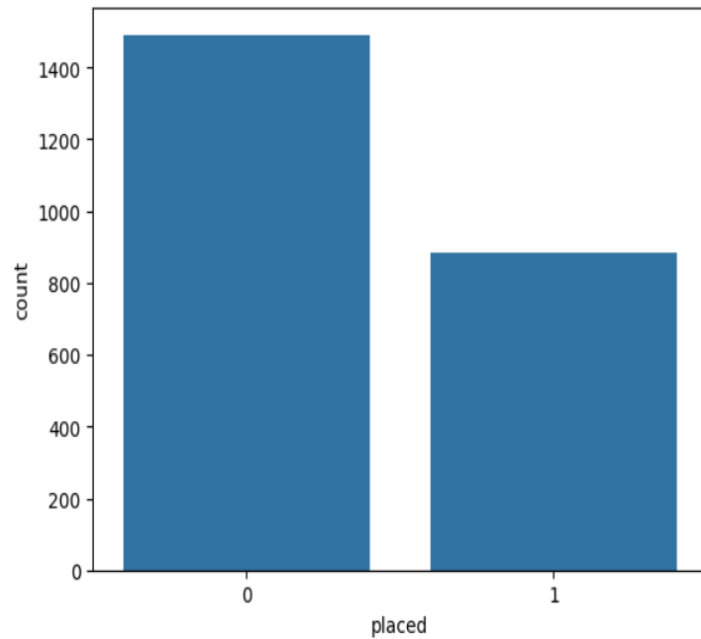
Program durations <15 days are excluded

2544

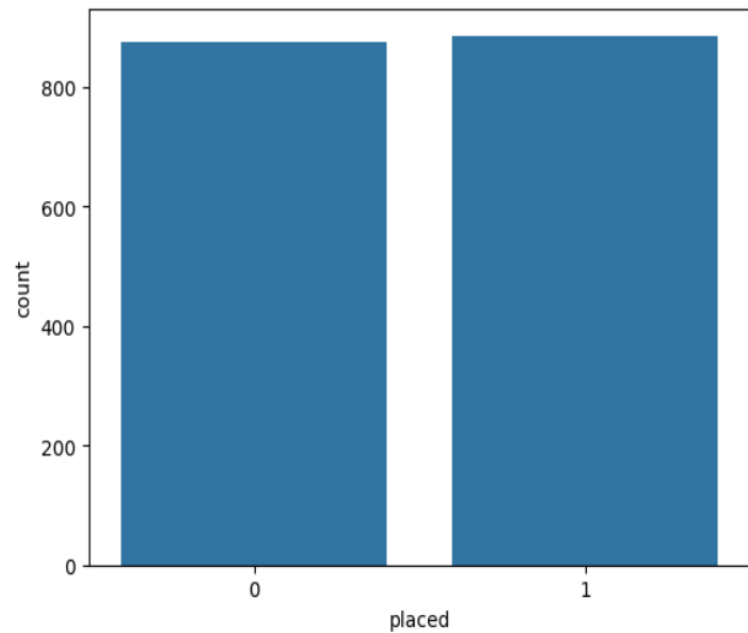


1894

Unballenced placed



Ballenced placed



✓ 0s [421] df.isnull().sum()



| | 0 |
|-----------------------------|---|
| pathrise_status | 0 |
| primary_track | 0 |
| program_duration_days | 0 |
| placed | 0 |
| employment_status | 0 |
| highest_level_of_education | 0 |
| length_of_job_search | 0 |
| biggest_challenge_in_search | 0 |
| professional_experience | 0 |
| work_authorization_status | 0 |
| number_of_interviews | 0 |
| number_of_applications | 0 |
| gender | 0 |
| race | 0 |

dtype: int64

There is no Null value in the used dataset.

```
df.drop(columns=['Unnamed: 0'], inplace=True)
df.drop(columns=['id'], inplace=True)
df.drop(columns=['cohort_tag'], inplace=True)
df.drop(columns=['pathrise_status'], inplace=True)
df.drop(columns=['number_of_interviews'], inplace=True)
```

Columns cohort_tag, pathrise_status, number_of_interviews and id have no effect on our prediction so they were excluded.

As we need to have **non-object** values in our dataset, we converted the columns with object type to int.

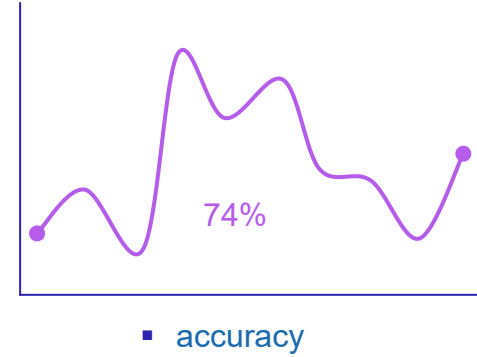
```
df=pd.get_dummies(df,columns=['employment_status ','primary_track','highest_level_of_education','length_of_job_search','gender','race','professional_experience'],dtype=int)
```



To automate tasks and gain insights from data,
we need to make a machine learning

model

- KNN is easy to understand and implement, as it relies on **calculating distances between data points** to make predictions without requiring complex training phases.
- KNN is a **non-parametric** algorithm which means it doesn't make any assume on structure of the underlying data.
- KNN **memorizes** training instances and uses them directly for predictions, which allows it to **adapt quickly to new data** but can make it slower and more memory-intensive for large datasets.





It was necessary to optimize

Hyperparameters

But what are they?

Hyperparameters necessity

In the KNN model, **hyperparameters** are essential settings that significantly affect the model's performance and accuracy. For example, the number of neighbors determines how many nearby points the model considers in making decisions. Choosing the right distance metric and weighting method helps the model achieve the best prediction results, especially with complex data.

- **Set Prior to Training:** Hyperparameters must be configured before the training process begins. They are not directly learned from the data but require manual selection and optimization.
- **Control Over Model Performance:** Hyperparameters play an essential role in managing the model's behavior and accuracy, helping to prevent issues like overfitting or underfitting.
- **Optimization Through Trial and Error:** Tuning hyperparameters is usually done through optimization techniques such as Grid Search or Random Search to achieve the best performance on a specific dataset.



which
Hyperparameters

do we used?

Used Hyperparameter

- ✓ `n_neighbors`: The number of neighbors in the K-Nearest Neighbors model, with values set between 1 and 10 to find the best value.



To optimize the hyperparameters, we have used

Grid Search

Grid Search

```
✓ [242] # Allows us to test parameters of classification algorithms and find the best one  
0s      from sklearn.model_selection import GridSearchCV
```

```
✓ [243] parameters = {'n_neighbors': [1, 2, 3, 4, 5, 6, 7, 8, 9, 10]}
```

```
✓ [244] knn_cv = GridSearchCV(knn_1, parameters)  
1s      knn_cv.fit(X_train, y_train)
```



What are the

Results

that we have obtained?

KNN Mean Absolute Error

- **MAE** calculates the average absolute difference between the predicted probabilities (usually between 0 and 1) and the actual labels (0 for benign, 1 for malware)
- Hyperparameter tuning reduced the model's average error by 0.5% (MAE).

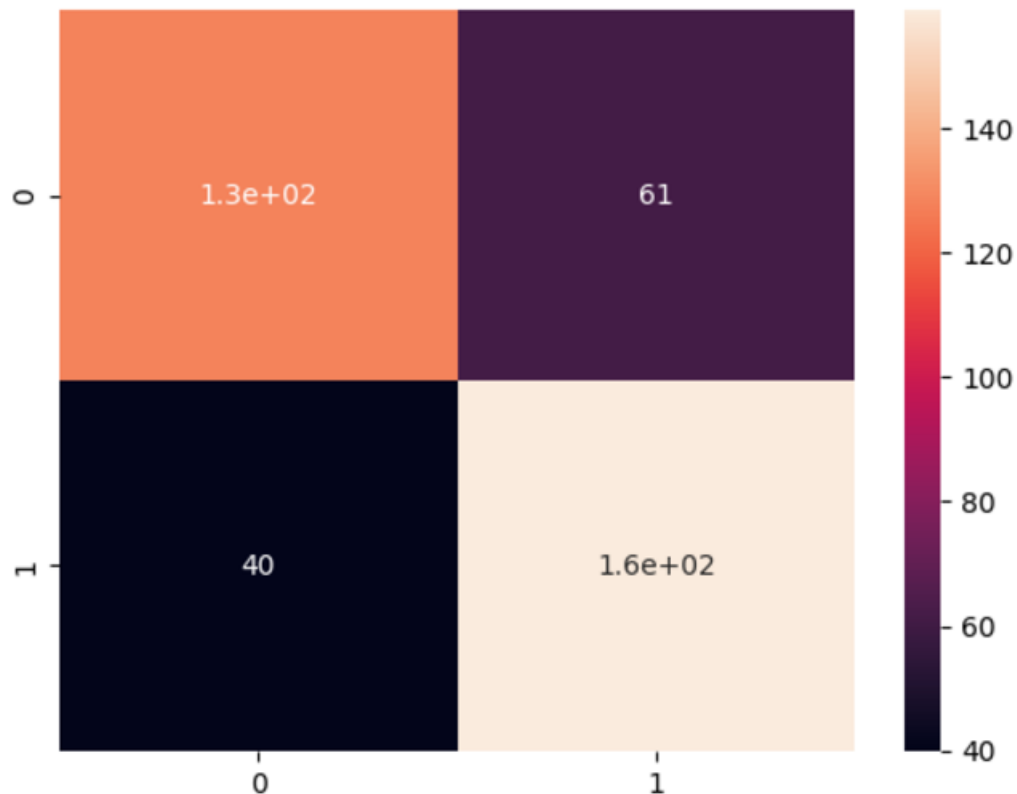
Before Grid Search : 26%

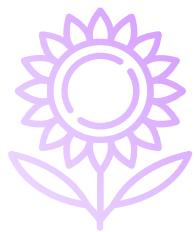
After Grid Search : 25.5%

▪ Mean Absolute Error

Confusion matrix

- We have used confusion matrix to visualize the performance of our classification model.
- It helps us to understand how many predictions the model got right and wrong for each class in the data.





THANKS

