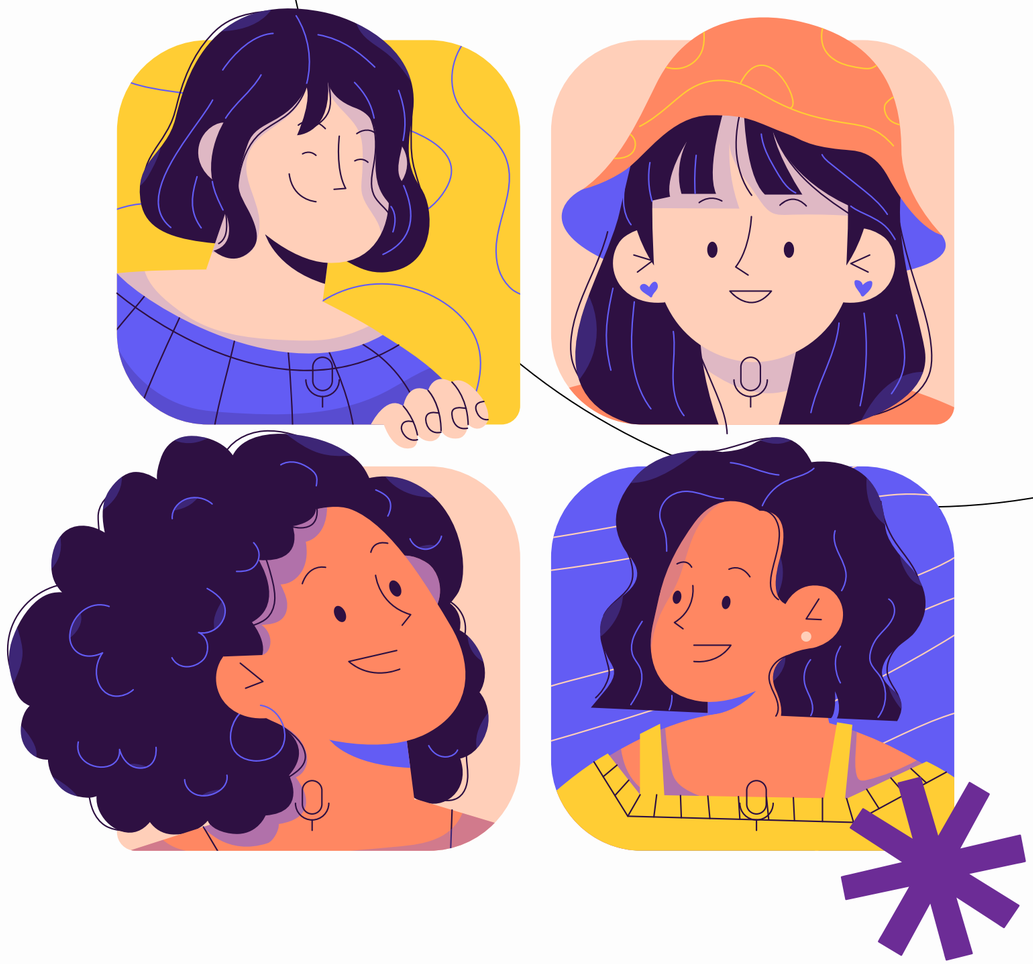
ALGORITMIKA PRO SCHOOL

SEEMS TO BE A HYPOTHESIS

BY KHANAPIN TAIR

With Teacher Vladislav Trufanov



A MINORITY OF GUYS GRADUATED IN A FOREIGN LANGUAGE

```
Number of guys with higher education: 299
Who know russian: 222
Other languages: 77
```

HOW I FOUND IT:

```python
def edu_3lvl_sex(row):
    global fem_phd, male_phd, lang_rus, lang_other
    if row['education_status'] == 3:
        if row['sex'] == 0:
            if row['langs'] == 1:
                lang_rus += 1
            else:
                lang_other += 1
            male_phd += 1
        else:
            fem_phd += 1
    return False
df['sex'] = df.apply(edu_3lvl_sex, axis = 1)

print('Number of guys with higher education:', male_phd)
print('Who know russian:', lang_rus)
print('Other languages:', lang_other)
```

this is just the final part*

MATHEMATICAL MODEL

```
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.neighbors import KNeighborsClassifier
from sklearn.metrics import confusion_matrix, accuracy_score

X = df.drop('result', axis = 1)
y = df['result']
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.25)

sc = StandardScaler()
X_train = sc.fit_transform(X_train)
X_test = sc.transform(X_test)
classifier = KNeighborsClassifier(n_neighbors = 5)
classifier.fit(X_train, y_train)

y_pred = classifier.predict(X_test)
print(y_test)
print(y_pred)
print('Percentage of correctly predicted outcomes:', round(accuracy_score(y_test, y_pred) * 100, 2))
print('Confusion matrix:')
print(confusion_matrix(y_test, y_pred))
```

ALWAYS COMES OUT DIFFERENT VALUE:

```
2751    1
4416    0
4147    0
348     1
5734    1
       ..
787     0
2781    1
5976    1
5246    0
1593    0
Name: result, Length: 2049, dtype: int64
[1 1 1 ... 1 1 1]
Percentage of correctly predicted outcomes: 57.0
Confusion matrix:
[[183 781]
 [100 985]]
```

```
6096    1
6462    0
7770    1
2950    0
3248    1
       ..
7064    0
6149    1
6116    0
5739    0
5903    0
Name: result, Length: 2049, dtype: int64
[1 0 0 ... 0 1 0]
Percentage of correctly predicted outcomes: 54.12
Confusion matrix:
[[392 567]
 [373 717]]
```

QUESTIONS CLARIFICATIONS?

Please feel free to contact me by email or phone.

Email Address

hanapin08@gmail.com

Mobile Number

+7-708-226-92-35

Consultation Hours

from 3 pm to 8 pm