



NAJAH

TO PREDICT THE SUCCESS OF YOUR BUSINESS

Machine Learning model

▶ **Phases to build our model**

- ▶ 1- collect our own data
- ▶ 2- preprocessing data (data cleaning – data analysis)
- ▶ 3- model selection
- ▶ 4- build model
- ▶ 5- test our model and display accuracy

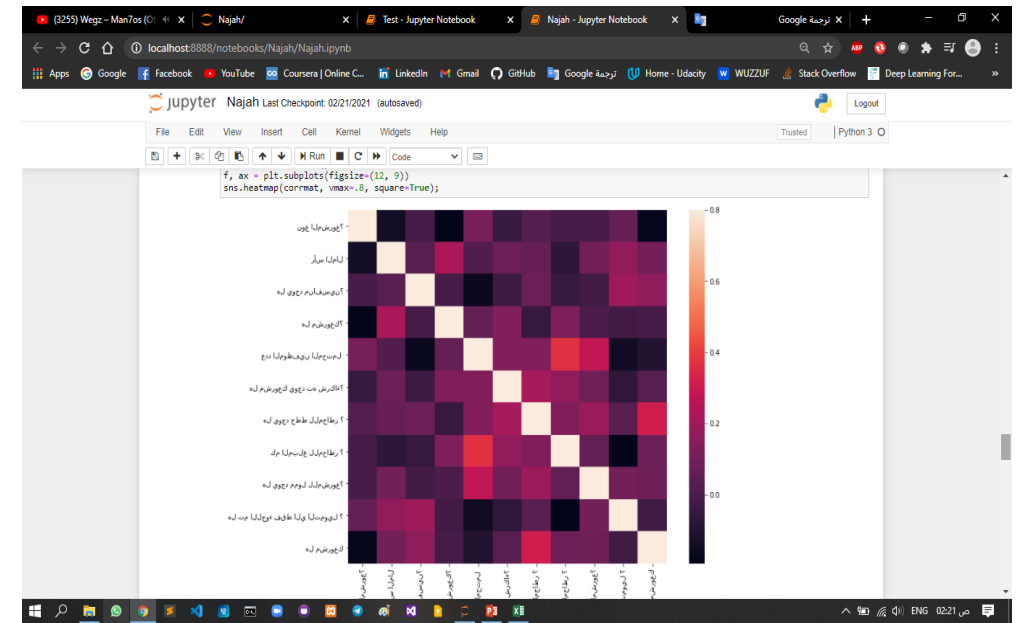
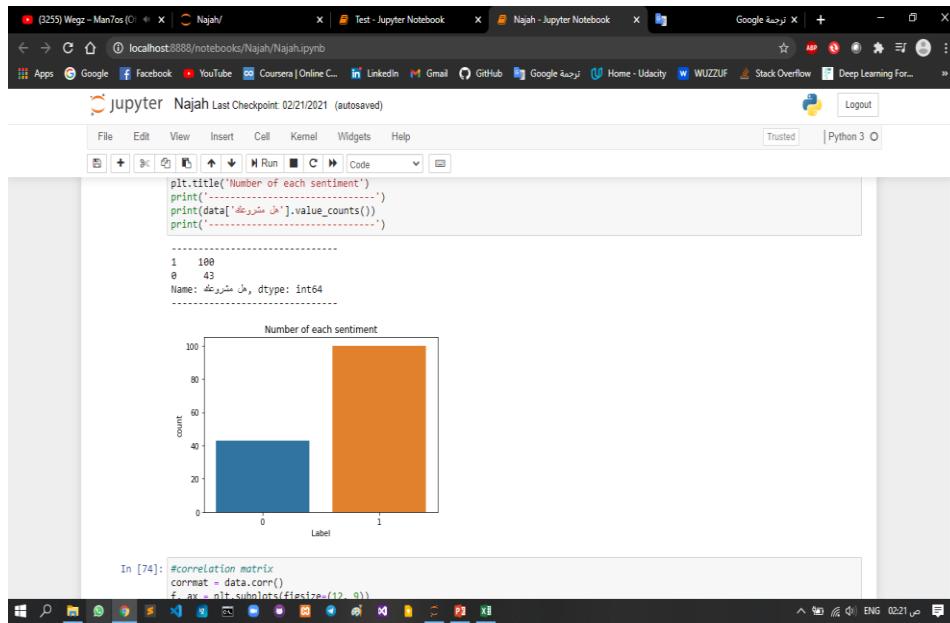
Collect Data

- ▶ We create google form that contains some questions that used as features in our model.
- ▶ Some example of our data:

| | A | B | C | D | E | F | G | H | I | J |
|-----|--------------------|--------------|----------------|------------------|-----------------|----------------------|--------------------------|----------------------|--------------------|---------------|
| 1 | Timestamp | نوع المشروع؟ | راس المال | هل يوجد منافسين؟ | هل مشروعك؟ | عدد الموظفين المحتمل | هل مشروعك يوجد به شركاء؟ | هل يوجد خطط للمخاطر؟ | كم المبلغ للمخاطر؟ | مودل للمشروع؟ |
| 113 | 2/15/2021 22:31:37 | معلم | 60000-100000 | نعم | عالي أرض الواقع | ١٠ و أقل | لا | نعم | أكثر من 1٥٠٠٠ | |
| 114 | 2/15/2021 22:39:20 | مكتبات | 10000-50000 | نعم | أون لاين | ١٠ و أقل | لا | نعم | 5000-10000 | |
| 115 | 2/15/2021 22:43:23 | مطاعم | 60000-100000 | نعم | أون لاين | ١٠ و أقل | لا | نعم | 10000-15000 | |
| 116 | 2/15/2021 22:57:12 | مطاعم | 10000-50000 | نعم | عالي أرض الواقع | ١٠ و أقل | لا | نعم | 5000-10000 | |
| 117 | 2/15/2021 23:11:36 | معلم | 10000-50000 | نعم | عالي أرض الواقع | ١٠ و أقل | لا | نعم | 5000-10000 | |
| 118 | 2/15/2021 23:23:37 | مطاعم | 60000-100000 | نعم | عالي أرض الواقع | ١٠ و أقل | نعم | نعم | 5000-10000 | |
| 119 | 2/15/2021 23:33:53 | معلم | 100000-5000000 | لا | عالي أرض الواقع | ٢٠ و أقل | نعم | لا | أكثر من 1٥٠٠٠ | |
| 120 | 2/15/2021 23:37:02 | ملاهي | 10000-50000 | نعم | عالي أرض الواقع | ١٠ و أقل | لا | نعم | أكثر من 1٥٠٠٠ | |
| 121 | 2/15/2021 23:41:24 | مطاعم | 10000-50000 | نعم | عالي أرض الواقع | ٢٠ و أقل | نعم | نعم | 5000-10000 | |
| 122 | 2/15/2021 23:45:18 | معلم | 10000-50000 | نعم | أون لاين | ١٠ و أقل | لا | نعم | 5000-10000 | |
| 123 | 2/16/2021 0:07:24 | مطاعم | 10000-50000 | نعم | عالي أرض الواقع | ١٠ و أقل | لا | نعم | 5000-10000 | |
| 124 | 2/16/2021 0:27:21 | مطاعم | 10000-50000 | نعم | عالي أرض الواقع | ١٠ و أقل | لا | نعم | 10000-15000 | |
| 125 | 2/16/2021 0:32:33 | مطاعم | 10000-50000 | نعم | أون لاين | ٢٠ و أقل | نعم | نعم | 5000-10000 | |
| 126 | 2/16/2021 1:28:10 | مطاعم | 10000-50000 | نعم | عالي أرض الواقع | ١٠ و أقل | لا | نعم | 5000-10000 | |
| 127 | 2/16/2021 1:38:57 | معلم | 100000-5000000 | نعم | عالي أرض الواقع | ٣٠ و أقل | نعم | نعم | أكثر من 1٥٠٠٠ | |
| 128 | 2/16/2021 2:05:58 | مطاعم | 10000-50000 | نعم | عالي أرض الواقع | ١٠ و أقل | لا | نعم | 5000-10000 | |
| 129 | 2/16/2021 2:13:39 | مطاعم | 60000-100000 | نعم | عالي أرض الواقع | ١٠ و أقل | نعم | نعم | 10000-15000 | |
| 130 | 2/16/2021 2:40:18 | مطاعم | 10000-50000 | نعم | أون لاين | ١٠ و أقل | لا | نعم | 5000-10000 | |
| 131 | 2/16/2021 2:58:58 | مطاعم | 10000-50000 | نعم | عالي أرض الواقع | ١٠ و أقل | نعم | نعم | 5000-10000 | |
| 132 | 2/16/2021 6:18:32 | مطاعم | 100000-5000000 | نعم | عالي أرض الواقع | ١٠ و أقل | لا | نعم | 10000-15000 | |
| 133 | 2/16/2021 8:39:42 | مطاعم | 10000-50000 | نعم | أون لاين | ١٠ و أقل | نعم | نعم | 5000-10000 | |
| 134 | 2/16/2021 10:23:35 | مطاعم | 10000-50000 | لا | عالي أرض الواقع | ١٠ و أقل | نعم | نعم | 5000-10000 | |
| 135 | 2/16/2021 13:43:05 | معلم | 100000-5000000 | نعم | عالي أرض الواقع | ٣٠ و أقل | نعم | نعم | أكثر من 1٥٠٠٠ | |
| 136 | 2/16/2021 18:48:14 | مطاعم | 10000-50000 | نعم | أون لاين | ١٠ و أقل | نعم | نعم | 5000-10000 | |
| 137 | 2/17/2021 21:07:59 | مطاعم | 10000-50000 | نعم | أون لاين | ١٠ و أقل | لا | نعم | 10000-15000 | |

Preprocessing

- ▶ Analysis our data and understand of its pattern to detect coloration of features
- ▶ Some plotting of data

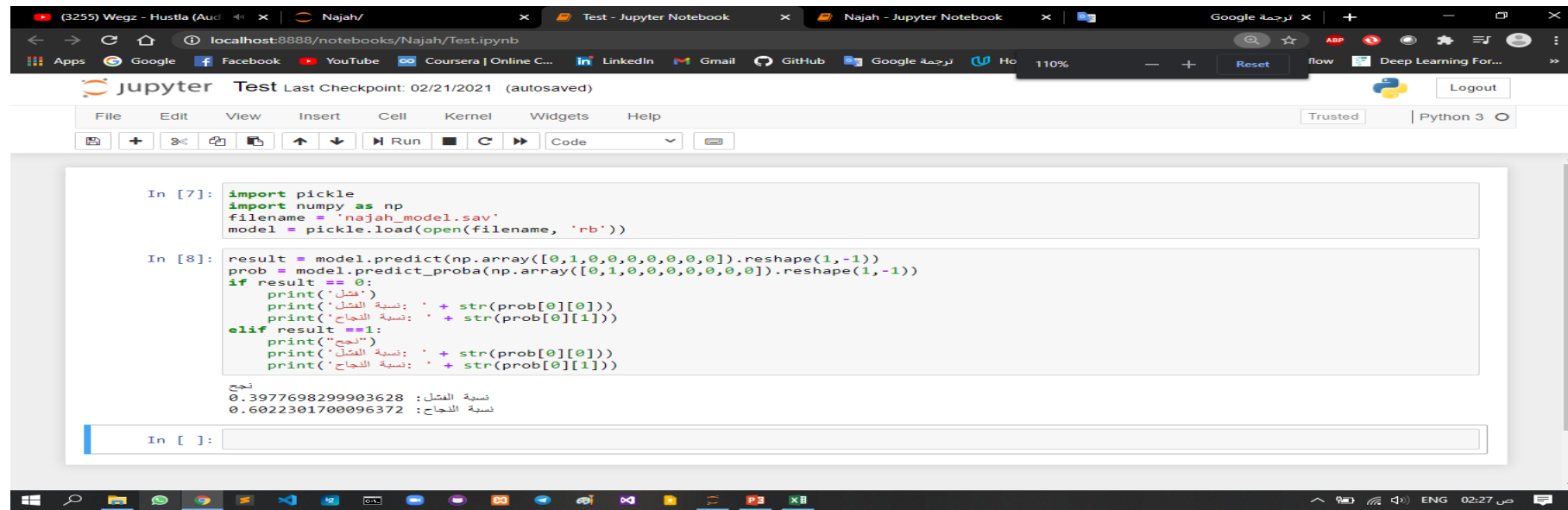


Model Selection

- ▶ Choose algorithm that fit our problem
- ▶ Algorithms chosen:
 - ▶ 1- Logistic Regression with accuracy : 79%
 - ▶ 2- SVM with accuracy: 76%
 - ▶ 3- Random forest with accuracy : 73%
 - ▶ 4- KNN with accuracy : 52%

Build Model

- ▶ 1- We choose logistic regression algorithm that give us the highest accuracy
- ▶ 2 – Then we test it with new data and get good result
- ▶ Example of text :



The screenshot shows a Jupyter Notebook interface with a browser window at the top. The notebook has a menu bar (File, Edit, View, Insert, Cell, Kernel, Widgets, Help) and a toolbar with icons for file operations, running code, and other functions. The code is written in Python and is divided into two input cells. The first cell imports the pickle module and loads a model from a file named 'najah_model.sav'. The second cell uses the loaded model to predict the class of a new data point and prints the result along with the predicted probabilities. The output of the second cell shows the model predicting class 0 (failure) with a probability of 0.3977698299903628 and class 1 (success) with a probability of 0.6022301700096372.

```
In [7]: import pickle
import numpy as np
filename = 'najah_model.sav'
model = pickle.load(open(filename, 'rb'))

In [8]: result = model.predict(np.array([0,1,0,0,0,0,0,0]).reshape(1,-1))
prob = model.predict_proba(np.array([0,1,0,0,0,0,0,0]).reshape(1,-1))
if result == 0:
    print('فشل')
    print('نسبة الفشل: ' + str(prob[0][0]))
    print('نسبة النجاح: ' + str(prob[0][1]))
elif result == 1:
    print('نجاح')
    print('نسبة الفشل: ' + str(prob[0][0]))
    print('نسبة النجاح: ' + str(prob[0][1]))

نجاح
0.3977698299903628 : نسبة الفشل
0.6022301700096372 : نسبة النجاح
```



Finish