Machine Learning -1

Semester Project

Presented by:
Amin Ghias - 25366
Uzair Zahidi - 26374

Problem Description

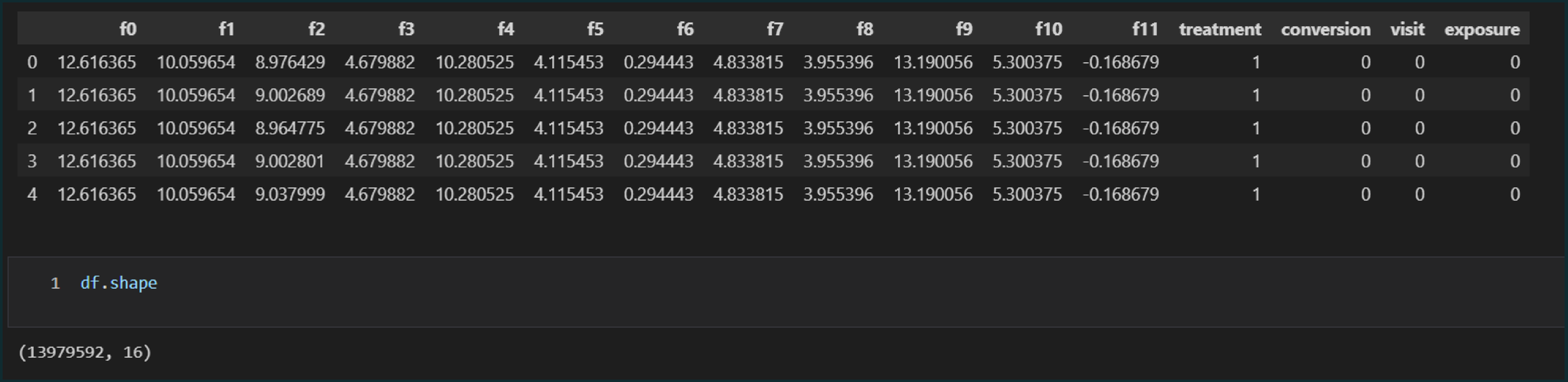
A large amount of money is spent by companies on advertisements. This expenditure on advertisement is costly and sometimes not very efficient. Companies are now trying to improve the efficiency of advertisements and target proper customers

This project will help in reducing the cost of advertisements and will pinpoint which people must be targetted for advertisement and which people not.

About the dataset

- The data was obtained from Kaggle "Uplift Modeling, Marketing and Campaign Data" provided by AI lab of Criteo
- The data contains 13 million instances from a randomized control trial collected in two weeks
- Each instance has 12 features that were anonymized plus a treatment variable and two target variables (visits and conversion).
- There is another extra variable called "exposure" which indicates whether the user was effectively exposed to the treatment

Initial dataset



	f0	f1	f2	f3	f4	f5	f6	f7	f8	f9	f10	f11	treatment	conversion	visit	exposure
0	12.616365	10.059654	8.976429	4.679882	10.280525	4.115453	0.294443	4.833815	3.955396	13.190056	5.300375	-0.168679	1	0	0	0
1	12.616365	10.059654	9.002689	4.679882	10.280525	4.115453	0.294443	4.833815	3.955396	13.190056	5.300375	-0.168679	1	0	0	0
2	12.616365	10.059654	8.964775	4.679882	10.280525	4.115453	0.294443	4.833815	3.955396	13.190056	5.300375	-0.168679	1	0	0	0
3	12.616365	10.059654	9.002801	4.679882	10.280525	4.115453	0.294443	4.833815	3.955396	13.190056	5.300375	-0.168679	1	0	0	0
4	12.616365	10.059654	9.037999	4.679882	10.280525	4.115453	0.294443	4.833815	3.955396	13.190056	5.300375	-0.168679	1	0	0	0

```python
df.shape
```

```
(13979592, 16)
```

Approach

Step 1: Divide the dataset into two parts on the basis of treatment

- Customers not being treated (treatment = 0)
- Customers being treated (treatment = 1)

Step 2: We will train our model-1 on dataset with treatment = 0 and predict on dataset with treatment = 1

Step 3: Dataset with treatment = 1 can be segmented into 4 classes

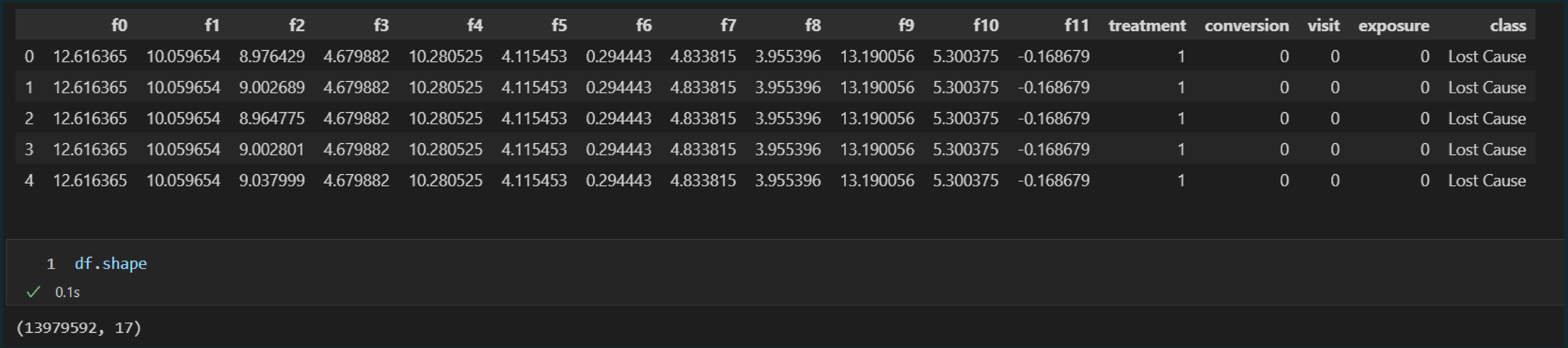
- Sure thing (Who will visit irrespective of treatment or not)
- Dont Disturb (Who will change form visiting to not visiting when treated with advertisment, that is negative effect of treatment)
- Lost Cause (Who will not visit irrespective of treatment or not)
- Persuadables (Who will visit when they are targetted)

Step 4: Now train our model-2 on treatment = 1 dataset and predict on dataset of treatment = 0

Step 5: Divide dataset of treatment = 0 on basis of predicted values into 4 smaller datasets of each class for treatment = 0

Step 6: Combine all the datasets with added class feature now our dataset has 4 classes

Final dataset for final model



	f0	f1	f2	f3	f4	f5	f6	f7	f8	f9	f10	f11	treatment	conversion	visit	exposure	class
0	12.616365	10.059654	8.976429	4.679882	10.280525	4.115453	0.294443	4.833815	3.955396	13.190056	5.300375	-0.168679	1	0	0	0	Lost Cause
1	12.616365	10.059654	9.002689	4.679882	10.280525	4.115453	0.294443	4.833815	3.955396	13.190056	5.300375	-0.168679	1	0	0	0	Lost Cause
2	12.616365	10.059654	8.964775	4.679882	10.280525	4.115453	0.294443	4.833815	3.955396	13.190056	5.300375	-0.168679	1	0	0	0	Lost Cause
3	12.616365	10.059654	9.002801	4.679882	10.280525	4.115453	0.294443	4.833815	3.955396	13.190056	5.300375	-0.168679	1	0	0	0	Lost Cause
4	12.616365	10.059654	9.037999	4.679882	10.280525	4.115453	0.294443	4.833815	3.955396	13.190056	5.300375	-0.168679	1	0	0	0	Lost Cause

```
df.shape
```

✓ 0.1s

```
(13979592, 17)
```

Step 7: Now Finally we will train a final-model on the new dataset with classes column to predict the class of customer.

In this way way we will tell apart based on predicted class if a customer needs advertisement treatment

Data Preprocessing

- Data Checked for null values
- Histogram of all features done
- Correlation matrix made

Models tried

- Logistic Regression
- Decission Tree
- Random Forrest
- Gradient Boosting
- Naive Bayes Gaussian
- K-NN
- Scaled k-NN
- Hist-Gradient Boosting

Evaluating Model performance

Model performance for finding final-model

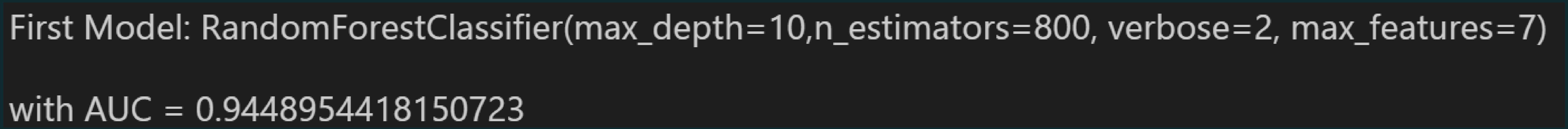
- Logistic : 0.9421809450740963
- Decision Tree: 0.9567119916019335
- Random Forest: 0.9626564673164548
-
- Gradient Boosting: 0.9633181808365909
- Hist-Gradient Boosting: 0.9651601462857793
- Naive Bayes Gaussian: 0.9339242229009771
- k-NN : 0.8172690444914528
- Scaled k-NN : 0.8600238102551588

Hyper parameter tuning

- Grid SearchCv used to find best parameters

Winner Models

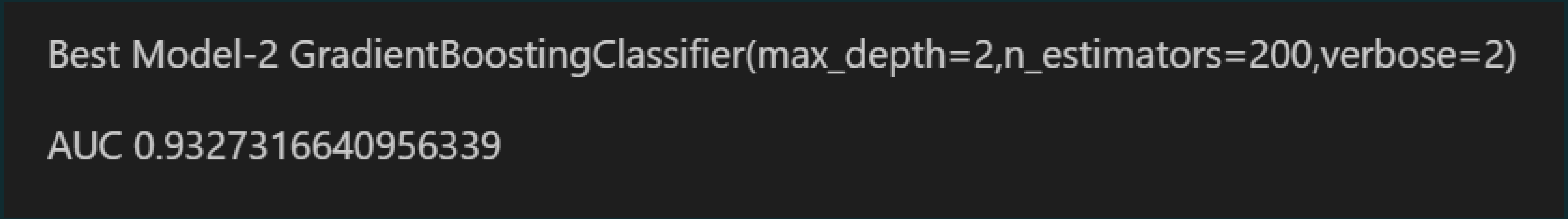
- Model-1



```
First Model: RandomForestClassifier(max_depth=10,n_estimators=800, verbose=2, max_features=7)

with AUC = 0.9448954418150723
```

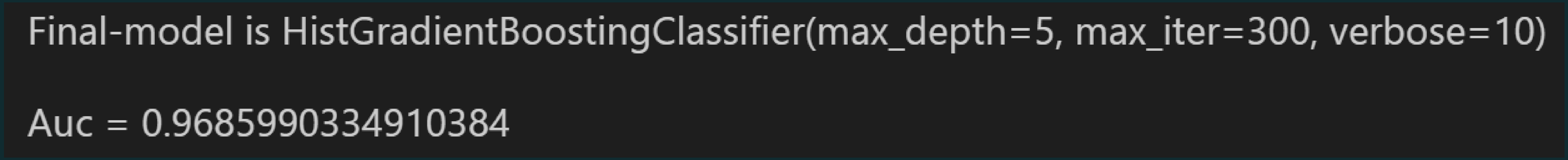
- Model-2



```
Best Model-2 GradientBoostingClassifier(max_depth=2,n_estimators=200,verbose=2)

AUC 0.9327316640956339
```

- Final-Model



```
Final-model is HistGradientBoostingClassifier(max_depth=5, max_iter=300, verbose=10)

Auc = 0.9685990334910384
```

Problems Faced

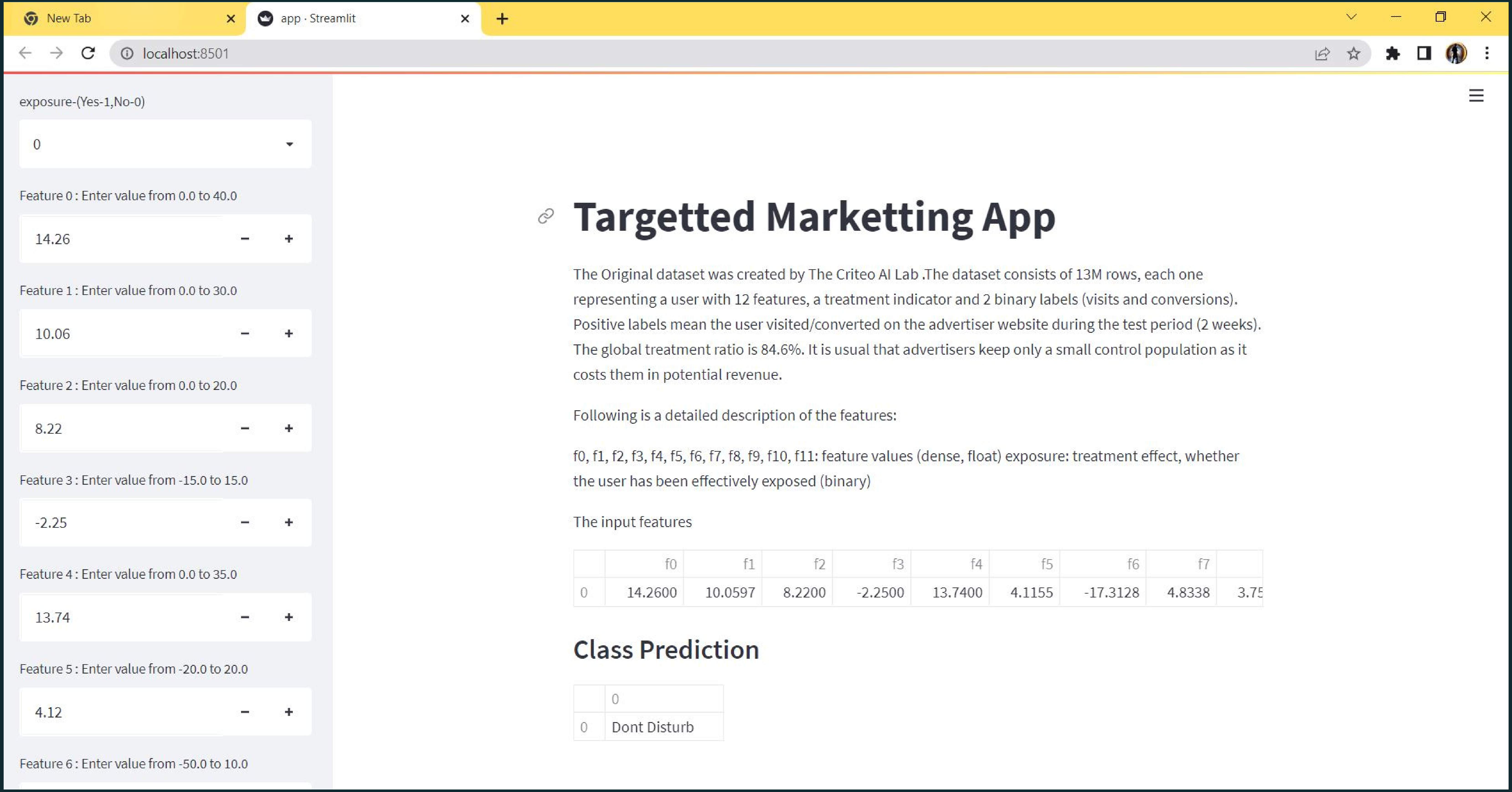
- The size of the dataset was very big so gridsearch could not be applied in the best way
- Finding the best parameters for our best model was time consuimg due to immense size of the dataset

Future Work

- More features can be added to the dataset to make better predictions
- This model can be modified and used by different sectors to redesign their advertisment methodology

Conclusion

- This project and the final model will help the marketting companies a lot, it will save their time and money and will enable them to target only the customers which need their product



exposure-(Yes-1,No-0)

0

Feature 0 : Enter value from 0.0 to 40.0

14.26

Feature 1 : Enter value from 0.0 to 30.0

10.06

Feature 2 : Enter value from 0.0 to 20.0

8.22

Feature 3 : Enter value from -15.0 to 15.0

-2.25

Feature 4 : Enter value from 0.0 to 35.0

13.74

Feature 5 : Enter value from -20.0 to 20.0

4.12

Feature 6 : Enter value from -50.0 to 10.0

Targetted Marketting App

The Original dataset was created by The Criteo AI Lab .The dataset consists of 13M rows, each one representing a user with 12 features, a treatment indicator and 2 binary labels (visits and conversions). Positive labels mean the user visited/converted on the advertiser website during the test period (2 weeks). The global treatment ratio is 84.6%. It is usual that advertisers keep only a small control population as it costs them in potential revenue.

Following is a detailed description of the features:

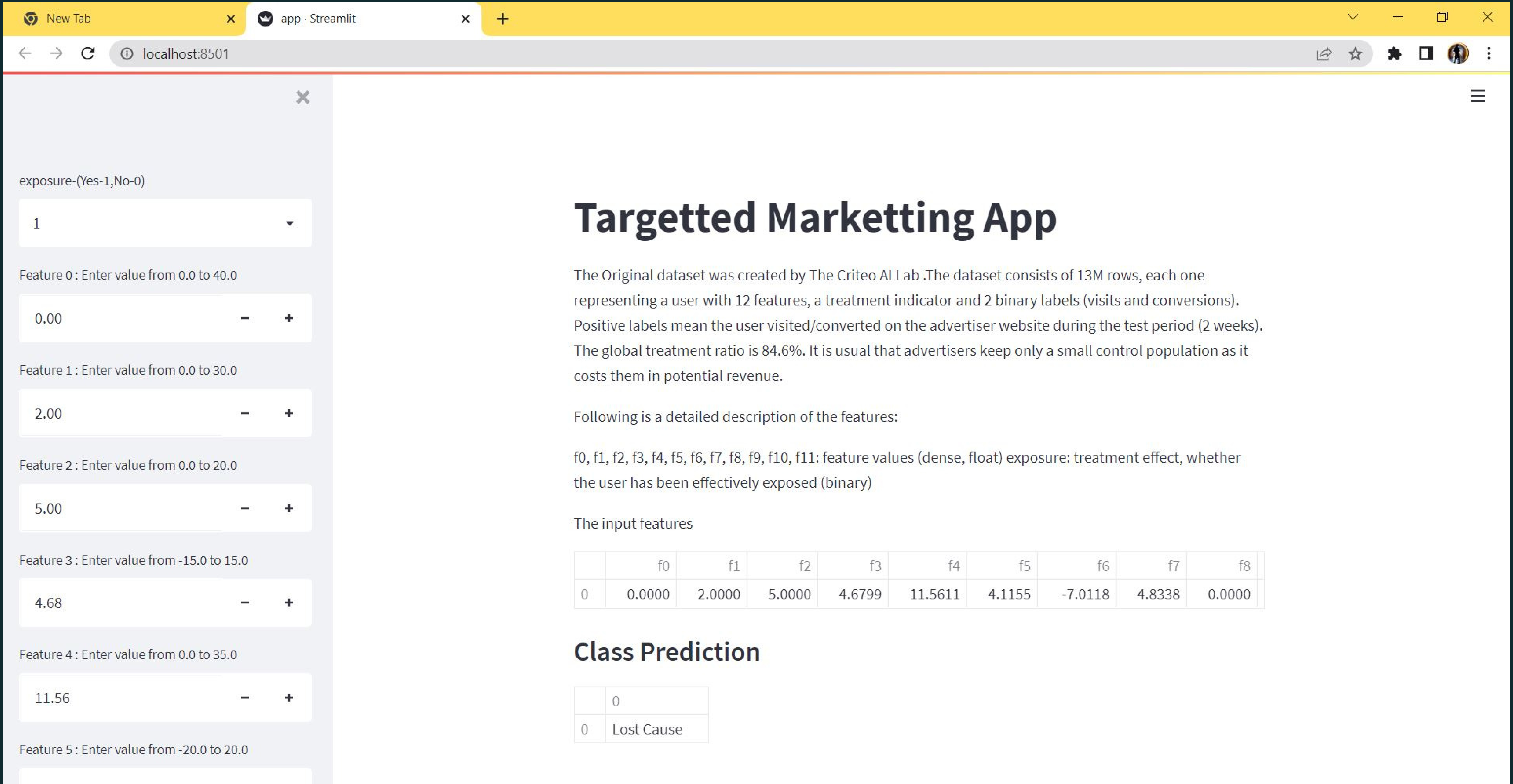
f0, f1, f2, f3, f4, f5, f6, f7, f8, f9, f10, f11: feature values (dense, float) exposure: treatment effect, whether the user has been effectively exposed (binary)

The input features

	f0	f1	f2	f3	f4	f5	f6	f7	
0	14.2600	10.0597	8.2200	-2.2500	13.7400	4.1155	-17.3128	4.8338	3.75

Class Prediction

	0
0	Dont Disturb



exposure-(Yes-1,No-0)

1

Feature 0 : Enter value from 0.0 to 40.0

0.00 − +

Feature 1 : Enter value from 0.0 to 30.0

2.00 − +

Feature 2 : Enter value from 0.0 to 20.0

5.00 − +

Feature 3 : Enter value from -15.0 to 15.0

4.68 − +

Feature 4 : Enter value from 0.0 to 35.0

11.56 − +

Feature 5 : Enter value from -20.0 to 20.0

Targetted Marketting App

The Original dataset was created by The Criteo AI Lab .The dataset consists of 13M rows, each one representing a user with 12 features, a treatment indicator and 2 binary labels (visits and conversions). Positive labels mean the user visited/converted on the advertiser website during the test period (2 weeks). The global treatment ratio is 84.6%. It is usual that advertisers keep only a small control population as it costs them in potential revenue.

Following is a detailed description of the features:

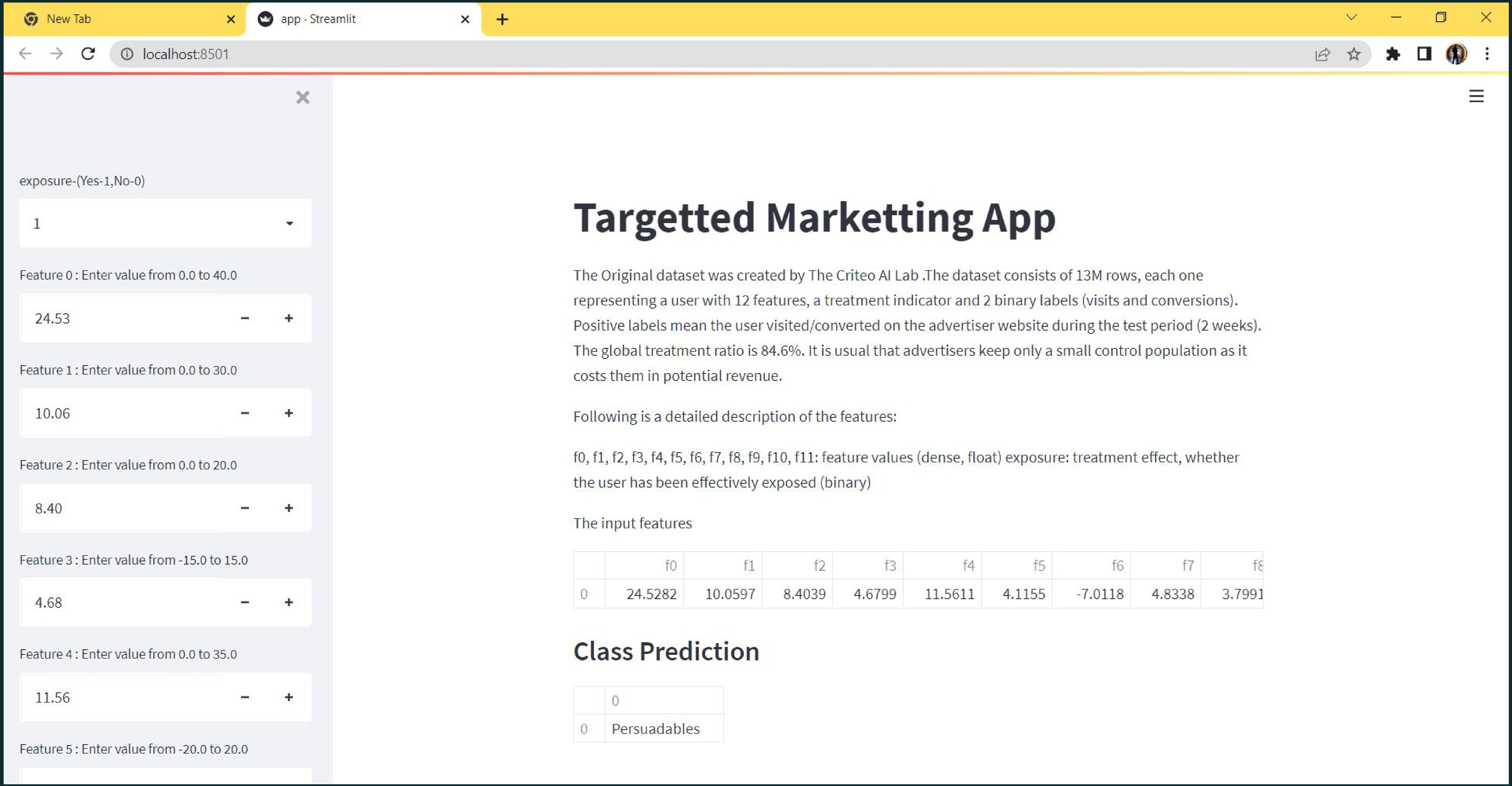
f0, f1, f2, f3, f4, f5, f6, f7, f8, f9, f10, f11: feature values (dense, float) exposure: treatment effect, whether the user has been effectively exposed (binary)

The input features

	f0	f1	f2	f3	f4	f5	f6	f7	f8
0	0.0000	2.0000	5.0000	4.6799	11.5611	4.1155	-7.0118	4.8338	0.0000

Class Prediction

	0
0	Lost Cause



exposure-(Yes-1,No-0)

1

Feature 0 : Enter value from 0.0 to 40.0

24.53

Feature 1 : Enter value from 0.0 to 30.0

10.06

Feature 2 : Enter value from 0.0 to 20.0

8.40

Feature 3 : Enter value from -15.0 to 15.0

4.68

Feature 4 : Enter value from 0.0 to 35.0

11.56

Feature 5 : Enter value from -20.0 to 20.0

Targetted Marketting App

The Original dataset was created by The Criteo AI Lab .The dataset consists of 13M rows, each one representing a user with 12 features, a treatment indicator and 2 binary labels (visits and conversions). Positive labels mean the user visited/converted on the advertiser website during the test period (2 weeks). The global treatment ratio is 84.6%. It is usual that advertisers keep only a small control population as it costs them in potential revenue.

Following is a detailed description of the features:

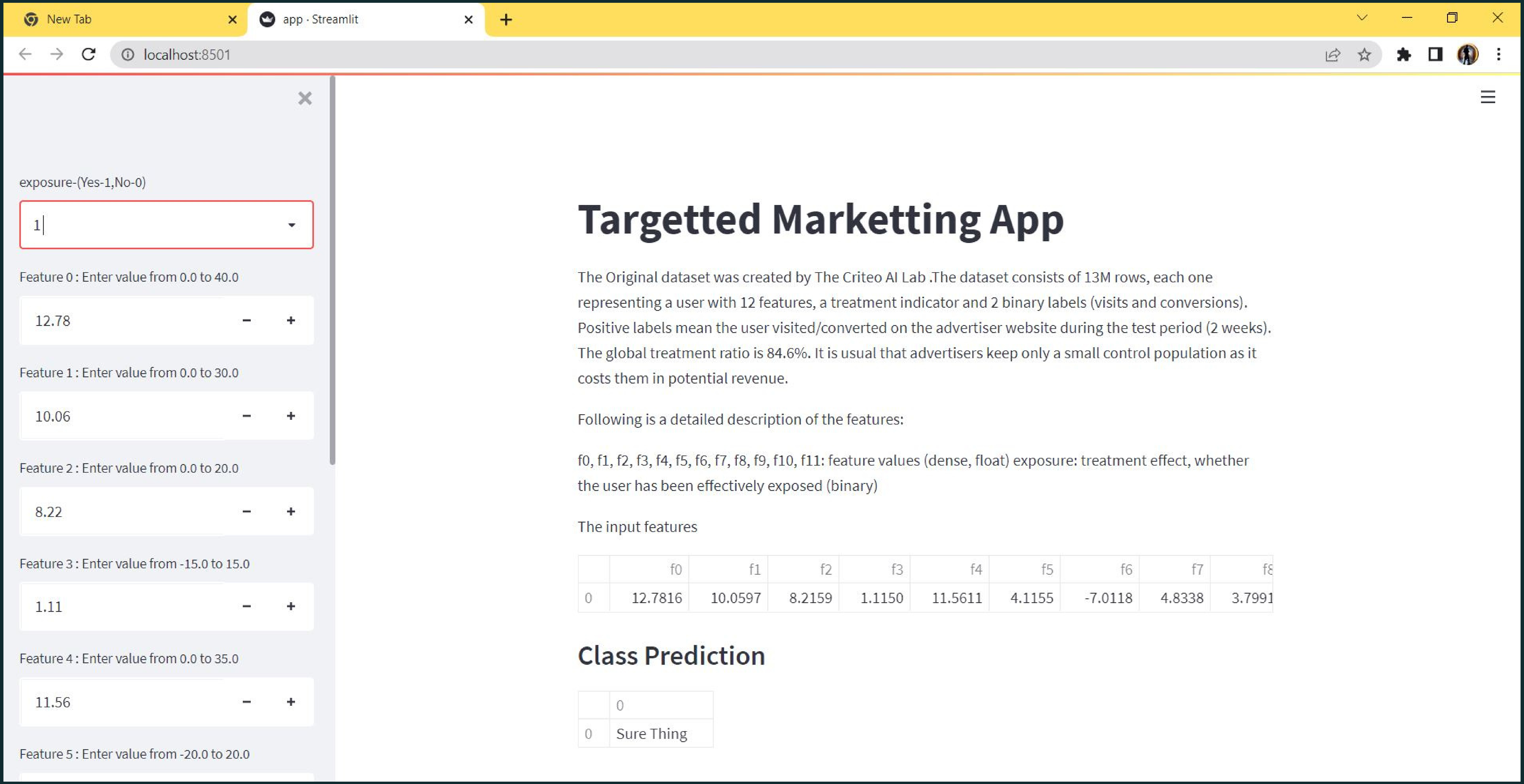
f0, f1, f2, f3, f4, f5, f6, f7, f8, f9, f10, f11: feature values (dense, float) exposure: treatment effect, whether the user has been effectively exposed (binary)

The input features

	f0	f1	f2	f3	f4	f5	f6	f7	f8
0	24.5282	10.0597	8.4039	4.6799	11.5611	4.1155	-7.0118	4.8338	3.7991

Class Prediction

	0
0	Persuadables



exposure-(Yes-1,No-0)

1

Feature 0 : Enter value from 0.0 to 40.0

12.78

Feature 1 : Enter value from 0.0 to 30.0

10.06

Feature 2 : Enter value from 0.0 to 20.0

8.22

Feature 3 : Enter value from -15.0 to 15.0

1.11

Feature 4 : Enter value from 0.0 to 35.0

11.56

Feature 5 : Enter value from -20.0 to 20.0

Targetted Marketting App

The Original dataset was created by The Criteo AI Lab .The dataset consists of 13M rows, each one representing a user with 12 features, a treatment indicator and 2 binary labels (visits and conversions). Positive labels mean the user visited/converted on the advertiser website during the test period (2 weeks). The global treatment ratio is 84.6%. It is usual that advertisers keep only a small control population as it costs them in potential revenue.

Following is a detailed description of the features:

f0, f1, f2, f3, f4, f5, f6, f7, f8, f9, f10, f11: feature values (dense, float) exposure: treatment effect, whether the user has been effectively exposed (binary)

The input features

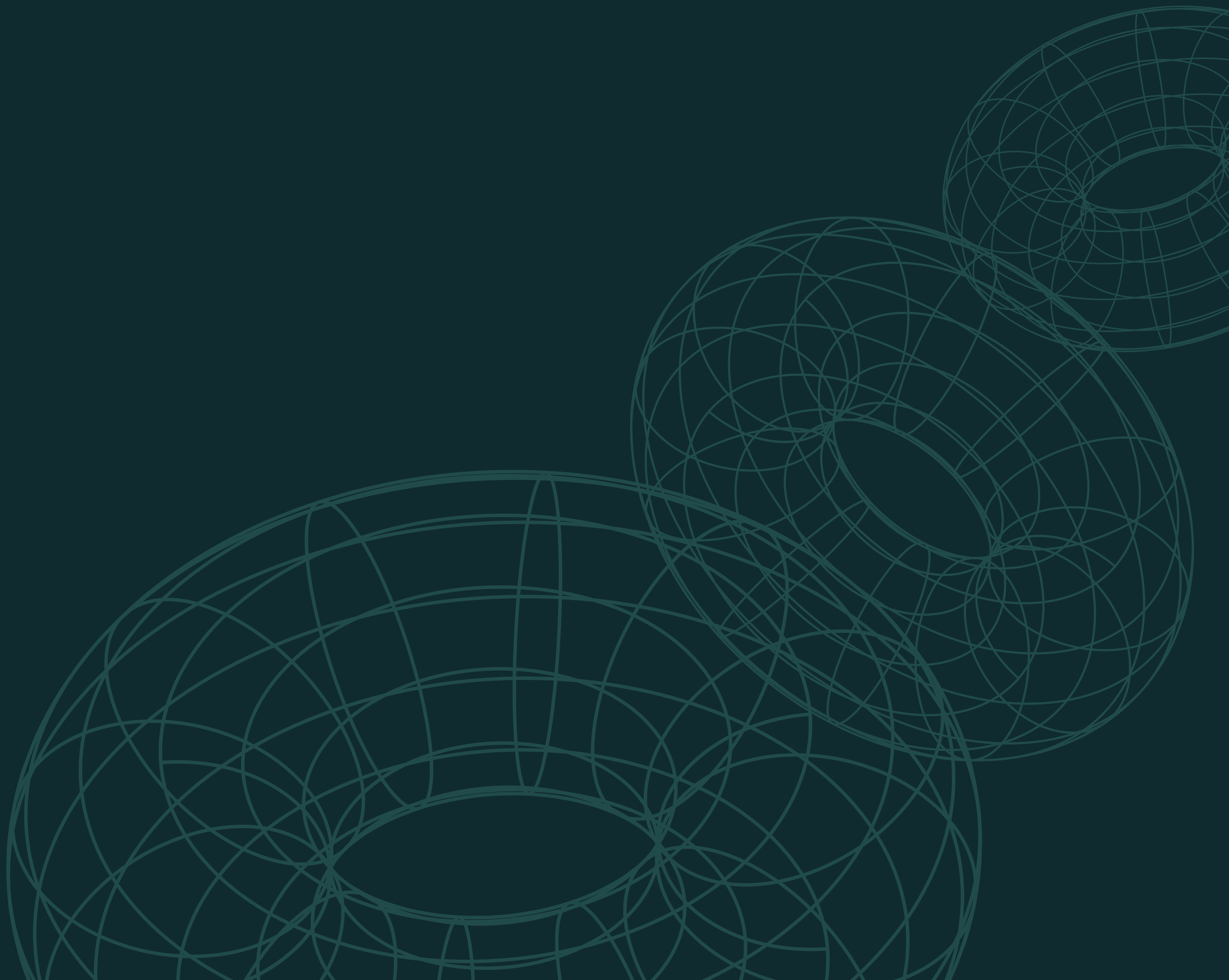
	f0	f1	f2	f3	f4	f5	f6	f7	f8
0	12.7816	10.0597	8.2159	1.1150	11.5611	4.1155	-7.0118	4.8338	3.7991

Class Prediction

	0
0	Sure Thing

Demo

Do you have any questions?



We hope you learned something new.