ML in PostgreSQL
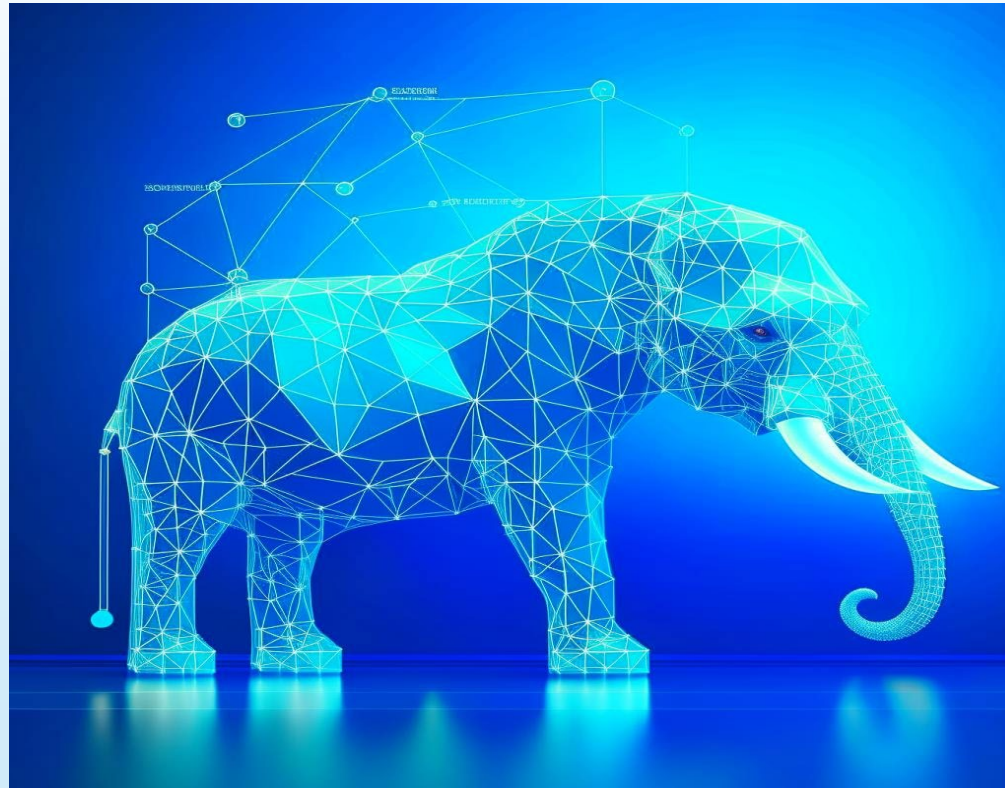
Using models:

- Binary Classification

- Multi Classification

- Regression

Using framework:

CatBoost: https://catboost.ai/

only prediction

ML process:

1. Training model
2. Save model to database

3. prediction


```
model.save_model('astra.cbm')
!ls
```

astra.cbm astra.json catboost_

```
adult=# SELECT ml_predict('astra3.cbm','astra3');
WARNING:  field run_id not used
WARNING:  field field_id not used
WARNING:  field spec_obj_id not used
WARNING:  field predict not used
      ml_predict
-----------------------
 public.astra3_predict
(1 row)
```

Installation

- git clone https://github.com/akalend/pg_ml.git
- export PG_HOME=/usr/local/pgsql   //where is main postgres folder
- wget https://github.com/catboost/catboost/releases/download/v1.2.2/libcatboostmodel.so
- mv libcatboostmodel.so $PG_HOME/lib
- cd pg_ml
- export PG_CONFIG=$PG_HOME/bin/pg_config
- export LD_LIBRARY_PATH=$PG_HOME/lib
- USE_PGXS=1 make
- sudo su
- export PATH=$PATH:$PG_HOME/bin
- USE_PGXS=1 make install
- chown postgres model.cbm
- [optional] cp model.cbm $PG_HOME/data

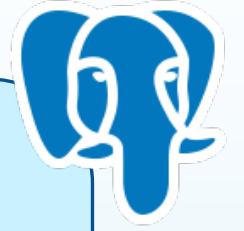# DataFrame columns as PostgreSQL fields

ID, Id, id

FieldID, Field_ID

Import-port

Port, PORT

<=>

id

field_id

import_port

port

DataFrame columns and PostgreSQL fields

```
[14] df = pd.read_csv('star_classification.csv')

[16] for it in df.columns:
         print(it)

     obj_ID
     alpha
     delta
     u
     g
     r
     i
     z
     run_ID
     rerun_ID
     cam_col
     field_ID
     spec_obj_ID
     class
     redshift
     plate
     MJD
     fiber_ID
```

<=>

```
adult=# \d astra3
                        Table "public.astra
    Column     |      Type         | Collatio
---------------+-------------------+----------
 alpha         | double precision  |
 delta         | double precision  |
 u             | double precision  |
 g             | double precision  |
 r             | double precision  |
 i             | double precision  |
 z             | double precision  |
 run_id        | bigint            |
 cam_col       | bigint            |
 field_id      | bigint            |
 spec_obj_id   | double precision  |
 redshift      | double precision  |
 plate         | bigint            |
 mjd           | bigint            |
 fiber_id      | bigint            |

adult=#
```

DataFrame columns and PostgreSQL fields

```
df.head()
```

	obj_ID	alpha	delta	u	g	r	i	z	run_ID	rerun_ID	cam_col	field_ID	spec_obj_ID
0	1.237661e+18	135.689107	32.494632	23.87882	22.27530	20.39501	19.16573	18.79371	3606	301	2	79	6.543777e+18
1	1.237665e+18	144.826101	31.274185	24.77759	22.83188	22.58444	21.16812	21.61427	4518	301	5	119	1.176014e+19
2	1.237661e+18	142.188790	35.582444	25.26307	22.66389	20.60976	19.34857	18.94827	3606	301	2	120	5.152200e+18
3	1.237663e+18	338.741038	-0.402828	22.13682	23.77656	21.61162	20.50454	19.25010	4192	301	3	214	1.030107e+19
4	1.237680e+18	345.282593	21.183866	19.43718	17.58028	16.49747	15.97711	15.54461	8102	301	3	137	6.891865e+18

```
adult=#  select * from astra3 limit 3;
     alpha        |       delta        |    u     |    g     |    r     |    i     |    z     | run_id | cam_col | field_id |      spec_obj_id      |
------------------+--------------------+----------+----------+----------+----------+----------+--------+---------+----------+-----------------------+
 16.95688978445004 | 3.64613008870454  | 23.33542 | 21.95143 | 20.48149 |   19.603 | 19.13094 |   7712 |       6 |      442 | 4.855016555329904e+18 |
 240.0632402477767 | 6.1341305981397  | 17.86033 | 16.79228 | 16.43001 | 16.30923 | 16.25873 |   3894 |       1 |      243 | 2.4489280322708705e+18 |
  30.887222067625 | 1.18870964120799  | 18.18911 | 16.89469 | 16.42161 | 16.24627 | 16.18549 |   7717 |       1 |      536 | 8.255357438959835e+18 |
(3 rows)

adult=#
```

Model information

```
adult=# SELECT ml_info ('astra3.cbm');
                                    ml_info
-----------------------------------------------------------------------------
 dimension:3 numeric features:12 categorial features:0 modelType "MultiClass"+
 fieldName:alpha,delta,u,g,r,i,z,cam_col,redshift,plate,MJD,fiber_ID
(1 row)

adult=#
```

```
adult=# SELECT ml_info ('titanic.cbm');
                                    ml_info
-----------------------------------------------------------------------------
 dimension:1 numeric features:2 categorial features:9 modelType "Accuracy"          +
 fieldName:PassengerId,Pclass,Name,Sex,Age,SibSp,Parch,Ticket,Fare,Cabin,Embarked
(1 row)
```

Model information

Model information

```
adult=# SELECT ml_info ('astra3.cbm');
                              ml_info
-----------------------------------------------------------------------------
 dimension:3 numeric features:12 categorial features:0 modelType "MultiClass"+
 fieldName:alpha,delta,u,g,r,i,z,cam_col,redshift,plate,MJD,fiber_ID
(1 row)

adult=#
```

- Dimension result
- Feature count (categorical and float)
- Type of model
- Fields name

Prediction of model

Path to model file

Table name

```
adult=# SELECT ml_predict ('astra3.cbm','astra3');
WARNING:  field run_id not used
WARNING:  field field_id not used
WARNING:  field spec_obj_id not used
WARNING:  field predict not used
      ml_predict
--------------------------
 public.astra3_predict
(1 row)
```

Create
the new table

Prediction results


```
adult=# SELECT * from astra3_predict;
 row |       alpha        |        delta         |    u     |    g     |    r     |    i     |    z     | run_id | cam_col | field_id |        spec_obj_id        |   redshift   | plate |  mjd  | fiber_id | predict  | class
-----+--------------------+----------------------+----------+----------+----------+----------+----------+--------+---------+----------+---------------------------+--------------+-------+-------+----------+----------+--------
   1 | 16.9568897845004   | 3.64613008870454     | 23.33542 | 21.95143 | 20.48149 | 19.603   | 19.13094 | 7712   | 6       | 442      | 4.855016555329904e+18     | 0.5062369    | 4312  | 55511 | 495      | 0.98686  | GALAXY
   2 | 240.063240247767   | 6.13413059813973     | 17.86033 | 16.79228 | 16.43001 | 16.30923 | 16.25873 | 3894   | 1       | 243      | 2.4489280322708705e+18    | 0.0003448142 | 2175  | 54612 | 348      | 0.990419 | STAR
   3 | 30.887222067625    | 1.18870964120799     | 18.18911 | 16.89469 | 16.42161 | 16.24627 | 16.18549 | 7717   | 1       | 536      | 8.255357438959835e+18     | 4.085216e-06 | 7332  | 56683 | 943      | 0.997588 | STAR
   4 | 247.594400505002   | 10.8877797153666     | 24.99961 | 21.71203 | 21.47148 | 21.30532 | 21.29109 | 5323   | 1       | 134      | 4.577998722756271e+18     | -0.0002914838| 4066  | 55444 | 336      | 0.997667 | STAR
   5 | 18.8964507920807   | -5.26133022886992    | 23.76648 | 21.79737 | 20.69543 | 20.23403 | 19.97464 | 7881   | 3       | 148      | 8.91047176642785e+18      | -0.0001361561| 7914  | 57331 | 303      | 0.996044 | STAR
   6 | 182.713733094955   | 51.3758050594777     | 22.44608 | 21.68444 | 20.24292 | 19.41423 | 19.08227 | 2830   | 1       | 411      | 7.516725588574623e+18     | 0.5026683    | 6676  | 56389 | 752      | 0.984373 | GALAXY
   7 | 150.089423193165   | 39.4670880748061     | 18.96441 | 17.82906 | 17.31429 | 16.99891 | 16.85583 | 3560   | 4       | 278      | 1.5267956411104236e+18    | 0.06366445   | 1356  | 53033 | 214      | 0.996164 | GALAXY
   8 | 189.510984338851   | 58.7411197772507     | 21.37376 | 20.80187 | 20.84925 | 21.13449 | 20.34689 | 2243   | 1       | 353      | 7.696817897528907e+18     | 0.7936153    | 6836  | 56443 | 684      | 0.957787 | QSO
   9 | 37.713872856097    | -0.525138228146508   | 20.77988 | 19.54618 | 19.16687 | 18.89438 | 18.64286 | 2700   | 2       | 117      | 1.7553283123029217e+18    | 0.1060118    | 1559  | 53271 | 183      | 0.993892 | GALAXY
  10 | 201.074980072746   | 28.7699058867715     | 25.05349 | 22.23362 | 20.8122  | 19.69488 | 19.28336 | 4649   | 3       | 120      | 7.306035245308205e+18     | 0.567082     | 6489  | 56329 | 247      | 0.993856 | GALAXY
  11 | 151.83091832672    | 19.8108624669417     | 24.04443 | 22.48608 | 20.59701 | 19.50985 | 19.00457 | 5183   | 5       | 142      | 6.622787444780849e+18     | 0.5475619    | 5882  | 56029 | 818      | 0.998885 | GALAXY
```

Prediction

Multi classification

```
adult=# SELECT * from astra3_predict;
 row |      alpha       |      delta        |   u     |    g    |   r     |   i     |   z     | run_id | cam_col | field_id |      spec_obj_id       |   redshift    | plate | mjd   | fiber_id | predict  | class
-----+------------------+-------------------+---------+---------+---------+---------+---------+--------+---------+----------+------------------------+---------------+-------+-------+----------+----------+--------
   1 | 16.9568897845004 |  3.64613008870454 | 23.33542| 21.95143| 20.48149|  19.603 | 19.13094|  7712   |    6    |   442    | 4.855016555329904e+18  |   0.5062369   | 4312  | 55511 |    495   | 0.98686  | GALAXY
   2 | 240.063240247767 |  6.13413059813973 | 17.86033| 16.79228| 16.43001| 16.30923| 16.25873|  3894   |    1    |   243    | 2.4489280322708705e+18 |  0.0003448142 | 2175  | 54612 |    348   | 0.990419 | STAR
   3 | 30.887222067625  |  1.18870964120799 | 18.18911| 16.89469| 16.42161| 16.24627| 16.18549|  7717   |    1    |   536    | 8.255357438959835e+18  |  4.085216e-06 | 7332  | 56683 |    943   | 0.997588 | STAR
   4 | 247.594400505002 | 10.8877797153666  | 24.99961| 21.71203| 21.47148| 21.30532| 21.29109|  5323   |    1    |   134    | 4.577998722756271e+18  | -0.0002914838 | 4066  | 55444 |    326   | 0.997667 | STAR
   5 | 18.8964507920807 | -5.26133022886992 | 23.76648| 21.79737| 20.69543| 20.23403| 19.97464|  7881   |    3    |   148    | 8.91047176642785e+18   | -0.0001361561 | 7914  | 57331 |    363   | 0.996044 | STAR
   6 | 182.71373094955  | 51.3758050594777  | 22.44608| 21.68444| 20.24292| 19.41423| 19.08227|  2830   |    1    |   411    | 7.516725588574623e+18  |   0.5026683   | 6676  | 56389 |    792   | 0.984373 | GALAXY
   7 | 150.089423193165 | 39.4670880748061  | 18.96441| 17.82906| 17.31429| 16.99891| 16.85583|  3560   |    4    |   278    | 1.5267956411104236e+18 |  0.06366445   | 1356  | 53033 |    274   | 0.996164 | GALAXY
   8 | 189.510984338851 | 58.7411197772507  | 21.37376| 20.80187| 20.84925| 21.13449| 20.34689|  2243   |    1    |   353    | 7.696817897528907e+18  |   0.7936153   | 6836  | 56443 |    604   | 0.957787 | QSO
   9 | 37.7138728560977 | -0.525138228146508| 20.77988| 19.54618| 19.16687| 18.89438| 18.64286|  2700   |    2    |   117    | 1.7553283123029217e+18 |   0.1060118   | 1559  | 53271 |    183   | 0.993892 | GALAXY
  10 | 201.074980072746 | 28.7699058867715  | 25.05349| 22.23362|  20.8122 | 19.69488| 19.28336|  4649   |    3    |   120    | 7.306035245308205e+18  |   0.567082    | 6489  | 56329 |    257   | 0.993856 | GALAXY
  11 | 151.83091832672  | 19.8108624669417  | 24.04443| 22.48608| 20.59701| 19.50985| 19.00457|  5183   |    5    |   142    | 6.622787444780849e+18  |   0.5475619   | 5882  | 56029 |    888   | 0.998885 | GALAXY
```

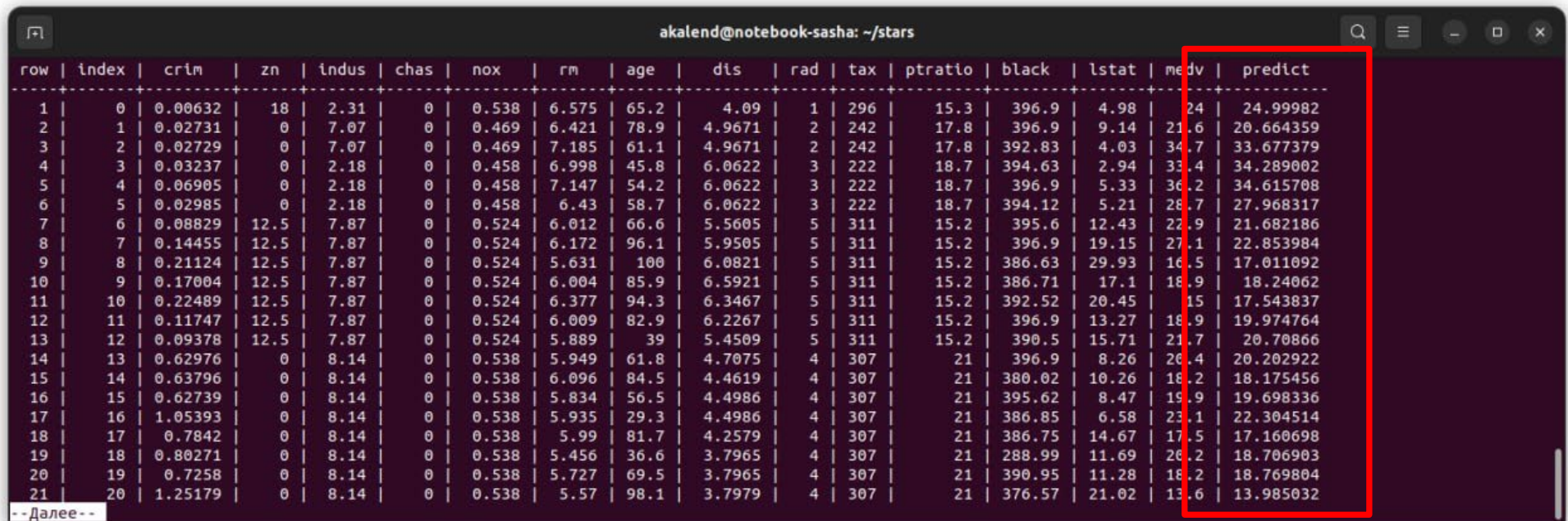## Binary classification

```
adult=# select * from titanic_predict;
 row | id  | passenger_id | pclass |                name                      |   sex  |  age  | sibsp | parch |  ticket   |  fare   | cabin | embarked | res | predict
-----+-----+--------------+--------+------------------------------------------+--------+-------+-------+-------+-----------+---------+-------+----------+-----+----------
 335 | 334 |     1226     |   3    | Cor, Mr. Ivan                            | male   |  27   |   0   |   0   | 349229    | 7.8958  | -999  |    S     |  f  | 0.115471
   1 |  0  |      892     |   3    | Kelly, Mr. James                         | male   | 34.5  |   0   |   0   | 330911    | 7.8292  | -999  |    Q     |  f  | 0.142616
   2 |  1  |      893     |   3    | Wilkes, Mrs. James (Ellen Needs)         | female |  47   |   1   |   0   | 363272    |    7    | -999  |    S     |  f  | 0.310916
   3 |  2  |      894     |   2    | Myles, Mr. Thomas Francis                | male   |  62   |   0   |   0   | 240276    | 9.6875  | -999  |    Q     |  f  | 0.083718
   4 |  3  |      895     |   3    | Wirz, Mr. Albert                         | male   |  27   |   0   |   0   | 315154    | 8.6625  | -999  |    S     |  f  | 0.125321
   5 |  4  |      896     |   3    | Hirvonen, Mrs. Alexander (Helga E Lindqvist) | female | 22 |   1   |   1   | 3101298   | 12.2875 | -999  |    S     |  f  | 0.39712
   6 |  5  |      897     |   3    | Svensson, Mr. Johan Cervin               | male   |  14   |   0   |   0   | 7538      |  9.225  | -999  |    S     |  f  | 0.112956
   7 |  6  |      898     |   3    | Connolly, Miss. Kate                     | female |  30   |   0   |   0   | 330972    | 7.6292  | -999  |    Q     |  t  | 0.643697
   8 |  7  |      899     |   2    | Caldwell, Mr. Albert Francis             | male   |  26   |   1   |   1   | 248738    |   29    | -999  |    S     |  f  | 0.253761
   9 |  8  |      900     |   3    | Abrahim, Mrs. Joseph (Sophie Halaut Easu) | female | 18   |   0   |   0   | 2657      | 7.2292  | -999  |    C     |  t  | 0.662787
  10 |  9  |      901     |   3    | Davies, Mr. John Samuel                  | male   |  21   |   2   |   0   | A/4 48871 |  24.15  | -999  |    S     |  f  | 0.037293
  11 | 10  |      902     |   3    | Ilieff, Mr. Ylio                         | male   | -999  |   0   |   0   | 349220    | 7.8958  | -999  |    S     |  f  | 0.093238
  12 | 11  |      903     |   1    | Jones, Mr. Charles Cresson               | male   |  46   |   0   |   0   | 694       |   26    | -999  |    S     |  f  | 0.210338
```

Prediction

Regression