

1. 5 колони, 150 редици, дрво на одлука треба да се направи, кој е најлошиот случај (worst case) максималниот број на листови?
2. Confusion matrix, class 1 (позитивна +), class 0 (негативна -), precision изнесува: 0.76, 0.87, 0.73, 0.68; Имаше дадено y оска со класите и x оска со некои други параметри во различен интензитет и бои... Heatmap изгледа.
3. Ако y временска серија имаме недостиг на податоци, на крајот (хронолошки), можат да се пополнат со: linear interpolation, forward fill, most frequent, backward fill...
4. Хиерархиско кластерирање, linkage критериум, доколку е дадена следната структура на кластери кој linkage не бил искористен? Complete linkage, ward linkage, single linkage, average linkage... Имаше плави точки во еден простор, потоа црвени во свој простор подалеку од овие и жолти во друг простор со многу точки...
5. 80 000 трансакции успешни, 75 000 неуспешни, 15 000 малициозни, со 90% точност...???
6. Пирсонова корелација (позитивна, силна позитивна, силна негативна, нема Пирсонова корелација...)... x и y оска и точки подредени во линеарна линија или дијагонално кон десно горе или од лево горе паѓаат надолу, кај мене беше почнувајќи од лево горе па се до десно доле...
7. Нормална дистрибуција, табела со Boston, New York и некои NULL вредности, значи ако ако се користи нормална дистрибуција што ќе биде следната торка или што ќе се предвиди следно...
8. Transformer модели, новитет кај себе имаат: convolution, self-attention, recurrent, pooling
9. Transformer architecture: encoder/decoder/encoder-decoder...
10. Ако користиме невронски мрежи за класифицирање на податоци во 10 различни класи, тогаш output слојот на невронската мрежа треба да има \_\_\_\_ (бројка) неврон/и и со \_\_\_\_ (на dropdown?) ат ф-ја.
11. Кога користиме регресија?
12. Кога користиме класификација?
13. За кои податоци користиме енкодирање со OrdinalEncoder()?
14. Кои се хиперпараметри во XGBoost?
15. Што е RandomForest и кои се неговите предности?