

Университет ИТМО

Проект

по дисциплине «Визуализация и моделирование»

Автор: Логвинов Лев Анатольевич

Поток: 11.03.02

Группа: К3220

Факультет: ИКТ

Преподаватель: Чернышева А.В.

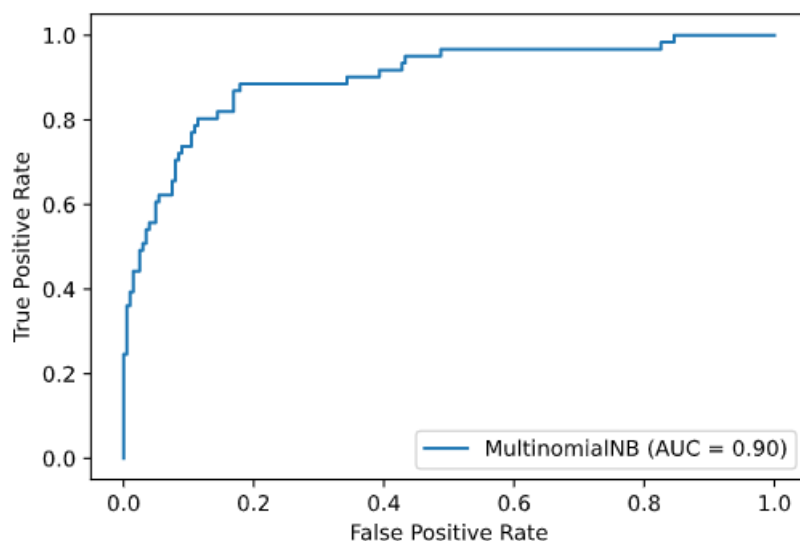
Санкт-Петербург, 2021 г.

Ссылка на датасет: [Датасет](#)

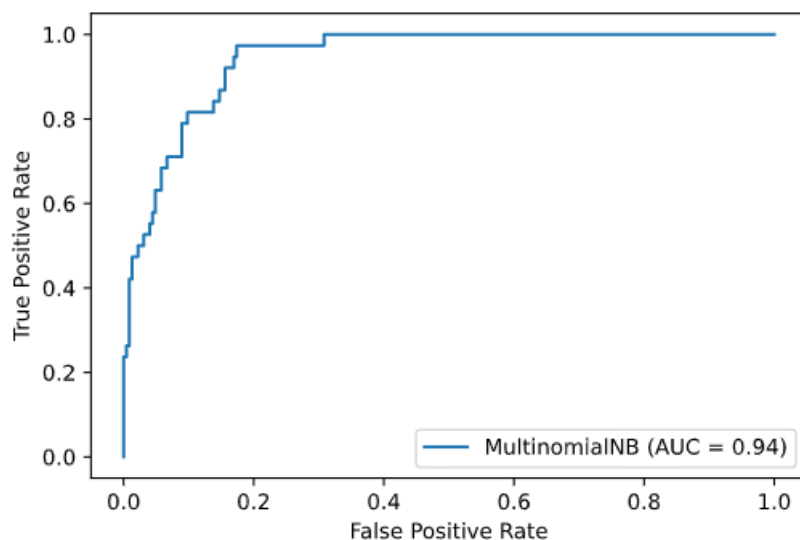
После обработки данный датасет содержит данные о погодных характеристиках в некотором городе и был ли там дождь в этот день.

Название столбца	Данные, хранящиеся в столбце	Тип данных
TempAvgF	Температура в градусах Фаренгейта	Целое число
DewPointAvgF	Точка росы в градусах Фаренгейта	Число с плавающей точкой
HumidityAvgPercent	Влажность в процентах	Число с плавающей точкой
SeaLevelPressureAvgInches	Давление на уровне моря в дюймах ртутного столбца	Число с плавающей точкой
VisibilityAvgMiles	Видимость в милях	Число с плавающей точкой
WindAvgMPH	Скорость ветра в милях в час	Число с плавающей точкой
PrecipitationSumInches	Общее количество осадков	Число с плавающей точкой
Events	Был дождь или нет	Целое число

Всего в датасете 1307 строк, 80% из которых мы зададим как тренировочные, а остальные 20% - как тестовые. Посмотрим, как отработает наивный байесовский классификатор на представленных данных. Получившаяся точность - 87%, является довольно неплохим результатом. Построим кривую рабочей характеристики приемника.



Теперь применим метод ближайших соседей для того же соотношения тренировочных и тестовых данных. Точность классификации при разных k примерно одинакова (около 80%), однако при $k=2$ является наибольшей и равна 84%. Построим кривую рабочей характеристики приемника.



Представим обе кривые на одном графике. Кривые ROC обычно показывают истинное положительное значение по оси Y и ложное срабатывание по оси X. Это означает, что верхний левый угол графика является «идеальной» точкой: ложноположительный показатель равен нулю, а истинный положительный показатель равен единице. Это означает, что большая площадь под кривой (AUC) обычно лучше. Как видно из графика, у нас получился хороший результат и для байесовского классификатора, и для метода KNN (при $k=2$).

