Прогнозирование оттока сотрудников компании (бинарная классификация)

При решении данной задачи вам необходимо проанализировать уход сотрудников из компании. Руководству компании необходимо взять данный процесс под контроль и суметь вовремя среагировать, чтобы минимизировать потери компании.

**Описание набора данных:**

**Целевая переменная.** Целевая переменная представляет собой флаг ухода сотрудника из компании.

**Описание признаков.** Датасет состоит из целевой переменной (Attrition), 34 признаков, включащих как числовые, так и категориальные признаки, всего в датасете - 1470 объектов.

*Age* - Возраст(Numeric)

*Attrition* - Флаг ухода из компании(String)

*BusinessTravel* - Частота командировок(String)

*DailyRate* - Ставка в день в долларах(Numeric)

*Department* - Департамент(String)

*DistanceFromHome* - Расстояние от дома до работы(Numeric)

*Education* - Образование(Numeric)

*EducationField* - Специальность(String)

*EmployeeCount* - Число подчиненных(Numeric)

*EmployeeNumber* - (Numeric)

*EnvironmentSatisfaction* - Удовлетворенность обстановкой в компании(Numeric)

*Gender* - Пол(String)

*HourlyRate* - Время работы(Numeric)

*JobInvolvement* - Вовлеченность в работу(Numeric)

*JobRole* - Должность(String)

*JobSatisfaction* - Удовлетворенность работй(Numeric)

*MaritalStatus* - Семейное положение(String)

*MonthlyIncome* - Доход в месяц(Numeric)

*MonthlyRate* - Ставка в месяц(Numeric)

*NumCompaniesWorked* - Число компаний, где работал сотрудник до текущей компании(Numeric)

*Over18* - Больше 18 лет(String)

*OverTime* - Сверхурочное время(String)

*PercentSalaryHike* - Надбавка к заработной плате(Numeric)

*PerformanceRating* - Рейтинг производительности(Numeric)

*RelationshipSatisfaction* - Удовлетворение отношением внутри коллектива(Numeric)

*StandardHours* - Стандартный график работы(Numeric)

*StockOptionLevel* - Сколько акций комании есть у сотрудника(Numeric)

*TotalWorkingYears* - Продолжительность трудового стажа(Numeric)

*TrainingTimesLastYear* - Общее время посещения тренировочных семинаров на работе(Numeric)

*WorkLifeBalance* - Соотношение свободного времени и работы(Numeric)

*YearsAtCompany* - Продолжительность работы в компании(Numeric)

*YearsInCurrentRole* - Продолжительность работы на текущей позиции(Numeric)

*YearsSinceLastPromotion* - Сколько прошло с последнего повышения(Numeric)

*YearsWithCurrManager* - Продолжительность работы с текущим менеджмером(Numeric)

**Задание:**

1. Изучить данные
2. Закодировать категориальные признаки;
3. Удалить признаки, являющиеся константами;
4. Выбрать признаки с высокой корреляцией, возможно, их стоит исключить из модели;
5. Разделить выборку на обучающую и тестовую
6. Оценить качество нескольких моделей.