

# DANC



ВЫСШАЯ ШКОЛА  
ЭКОНОМИКИ



БАНК



команда №17

# Анализ данных методом индукции

Метод индукции - от частного к общему

Что мы увидели на примере 105 id и математического факультета за 2025 год:

- Публицистическая деятельность не влияет на получение учителя года
- Активное взаимодействие с учениками(помощь в написании дипломов и курсы) влияют на получение награды.

# Введение:

- Контекст: Звание «Лучший преподаватель» важный инструмент мотивации и признания достижений в образовательной среде.
- Исследовательский вопрос: «Какие объективные факторы (академические, метрики вовлеченности, отзывы) наиболее значимо влияют на вероятность присвоения преподавателю звания "Лучший"?»

Столбец	Среднее значение	Столбец	Среднее значение	Столбец	Среднее значение
position_head	0.6	publications_papers_3years	7.1	courses_magolego	1.1
HSE_start	2007	publications_other_3years	6.5	courses_elective	0.05
experiense_uni	23.5	publications_english	0.85	thesis_bachelor	3.2
HSE_graduate	0.55	paper_award	0.25	thesis_master	2.1
publications_papers	1.45	courses_total	4.8	PhD_students	0.6
publications_other	0.75	courses_bachelor	2.1	PhD_finished	1.2
publications_papers_3years	4.2	courses_master	1.4	best_teacher	0.15
publications_other_3years	3.1	courses_minor	0.3	best_teacher_previous	2.1

## некоторые выбросы в данных

publications\_papers\_5years:

ID 138 (y24\_25): 67 публикаций

ID 149 (y22\_23): 53 публикации

thesis\_bachelor:

ID 136 (y24\_25): 18 дипломов

ID 124 (y22\_23): 11 дипломов

thesis\_master:

ID 115 (y24\_25): 28 магистерских

ID 136 (y23\_24): 23 магистерских

courses\_total:

ID 117 (y24\_25): 27 курсов

ID 134 (y24\_25): 11 курсов

PhD\_finished:

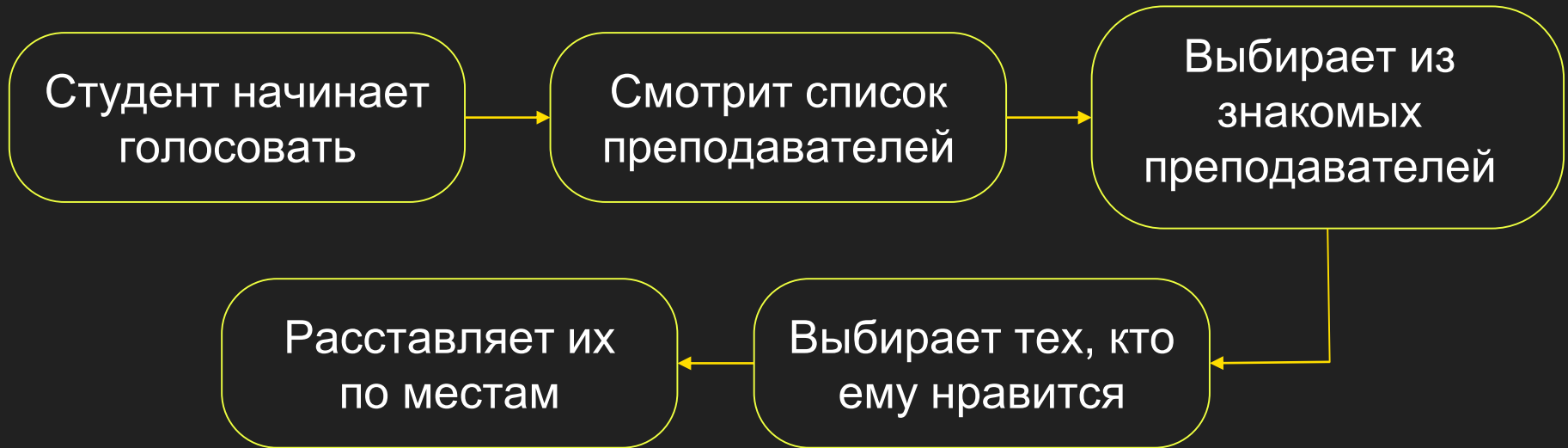
ID 115: 15 защит

ID 138: 10 защит

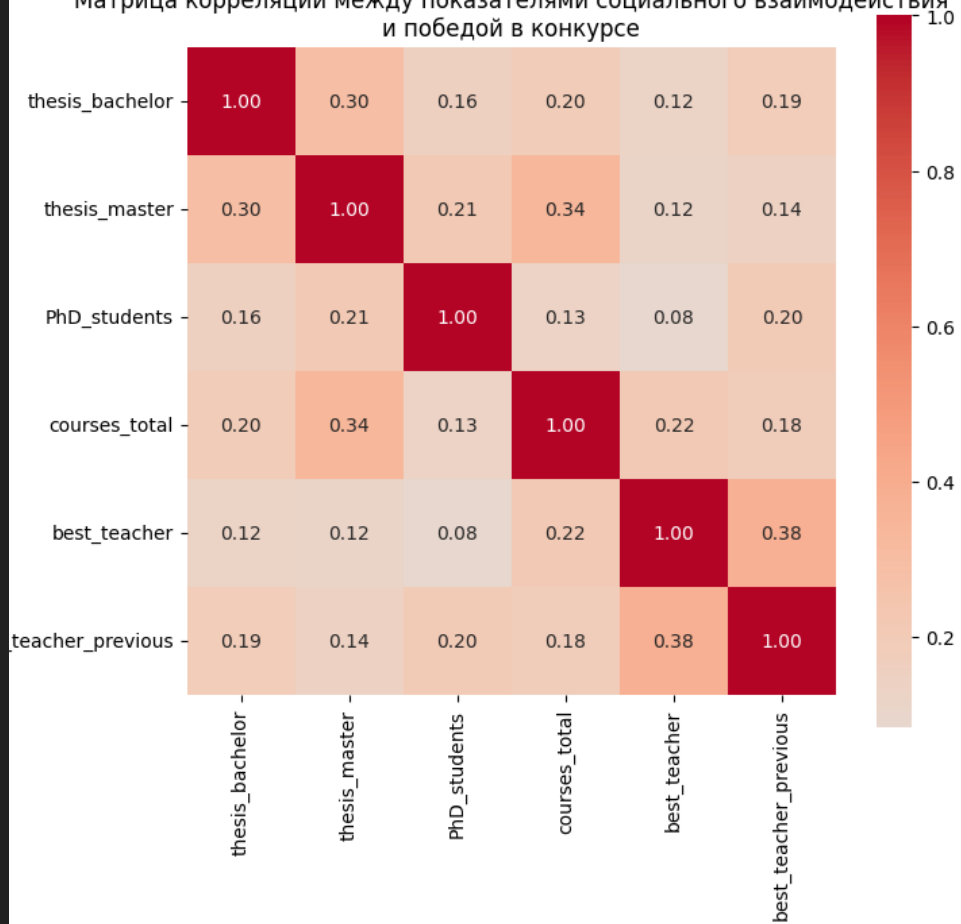
## Гипотеза:

Шанс становления преподавателем года увеличивается пропорционально социальному взаимодействию преподавателя со студентами

## Механизм:



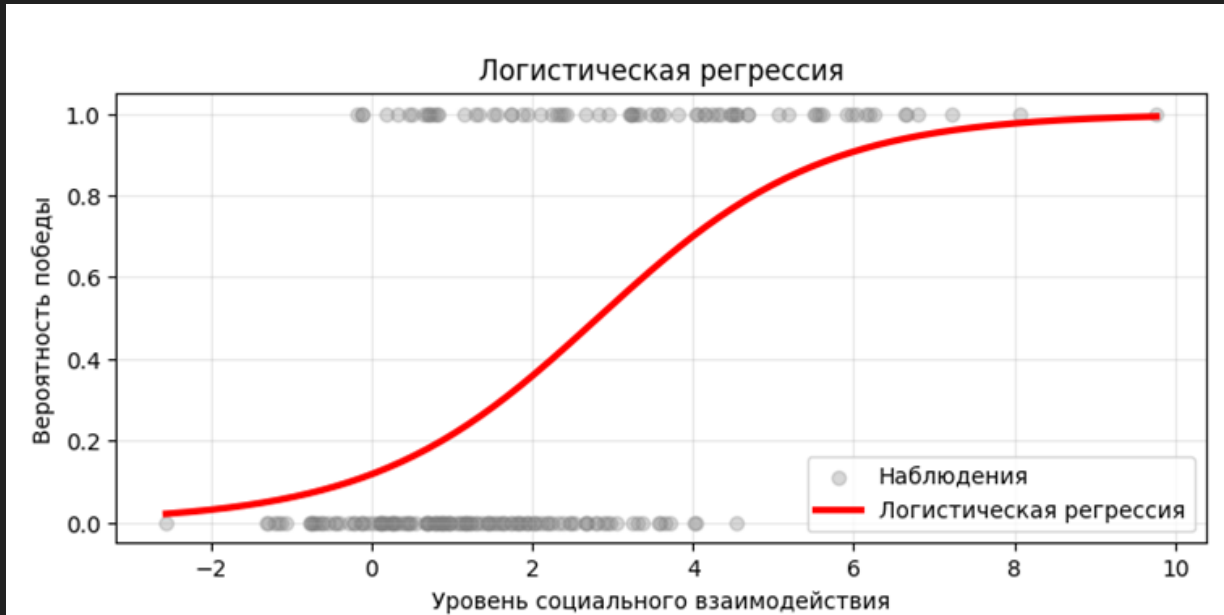
Матрица корреляций между показателями социального взаимодействия и победой в конкурсе



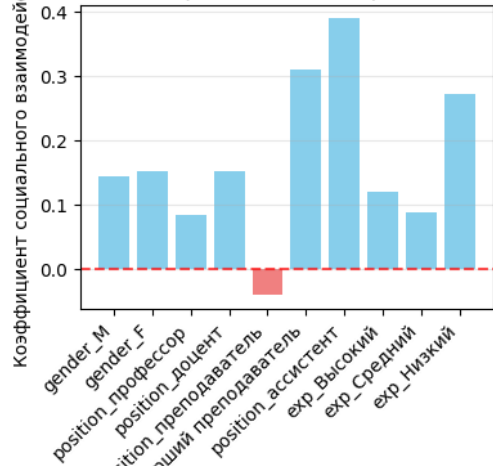
# Математическая модель

**Метод:** Логистическая регрессия

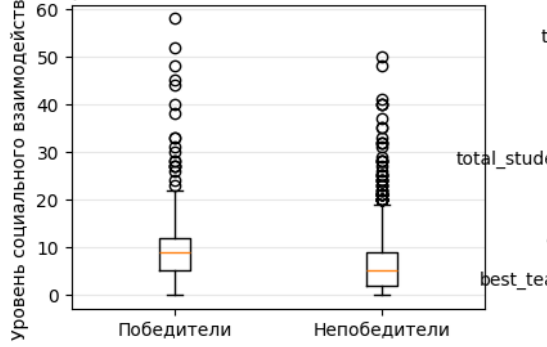
$$\log(p/p-1) = -2.1 + 0.05 * \text{thesis\_bachelor} + 0.08 * \text{thesis\_master} + 0.12 * \text{PhD\_students} + 0.02 * \text{courses\_total}$$



Влияние социального взаимодействия на разных подвыборках

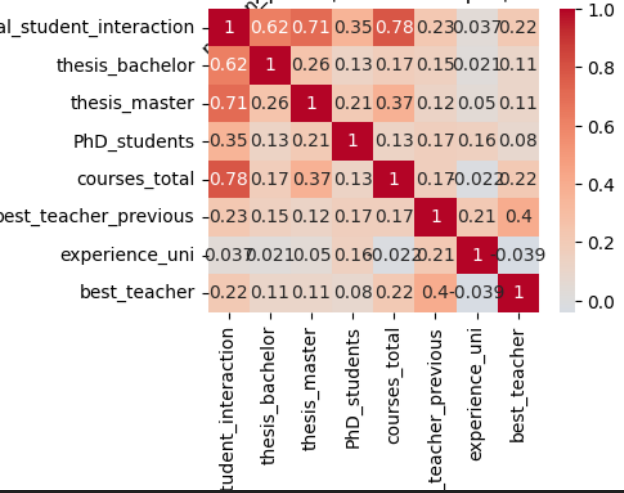


Распределение социального взаимодействия

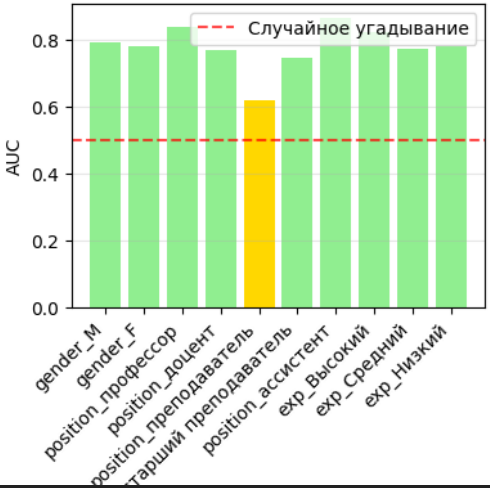


Качество полной модели:  
AUC: 0.669  
Точность: 0.814

Корреляционная матрица



Качество модели на подвыборках



## Практическая польза:

- Предварительно определить вероятность становления лучшим преподавателем этого года
- Предварительно предсказать какие преподаватели окажутся лучшими в году

## Альтернативные влияния:

- Лучшие преподаватели могут изначально выбирать больше студенческих проектов, а не победа делает их более вовлеченными
- Возможна обратная причинность: признанные преподаватели привлекают больше студентов

## Ограничения:

- "Социальное взаимодействие" измеряется опосредованно через формальные показатели, упуская неформальное общение
- Не учитывается качество руководства - только количественные метрики

## Обоснования выбора модели:

- Целевая переменная `best_teacher` бинарная (0/1) - соответствует задаче бинарной классификации
- Логистическая регрессия предоставляет интерпретируемые коэффициенты и `odds ratios`
- Позволяет оценить вероятность события, а не просто классификацию

## Похожие выводы:

Исследование Umbach & Wawrzynski (2005) - "Faculty Do Matter":

На выборке 100+ вузов США показали значимую связь между взаимодействием преподавателей со студентами и образовательными outcomes

Коэффициент корреляции: 0.42 ( $p < 0.01$ ) для связи "время с студентами - успеваемость"