Здравствуйте!

Я Дарья и сегодня хочу рассказать о том, что я сделала за это время по проекту построение модели времени дожития пациентов с БАС.

2) Как вам известно так как признаки водородных связей представлены донорно-акцепторным взаимодействием фермента с описывающим его весом, то для каждого пациента фермент рассматривался в виде графа.

3)Для каждого такого графа будем считать характеристики этого графа и на основе этих строить обучающую выборку. Мы использовали линейную регрессию, рассматриваемое уравнение вы можете видеть на слайде.

4) В качестве ретроспективы хочу напомнить какие этапы были сделаны, до сегодняшнего дня.

5) Во-первых это рассмотрение графа с различным представлением вершины 5, 6, 7

8) Для каждого из этих графов я считала представленные на слайде характеристики и строила модель.

9) **)** Построенные модели показали маленький коэффициент детерминации, что означает плохое описание данных. И как выяснилось это обусловлено тем, что Характеристики для всего графа не учитывают структуру конкретных подграфов, в которых может возникнуть мутация.

10)Поэтому сейчас все исследования ведутся на графе вершины которого представлены в виде полной водородной связи. Так как посчитанные характеристики для всего графа не дали нужных результатов, было принято решение посчитать характеристики для каждой вершины графа и на данном этапе возникла проблема вырожденности, обусловленная большим количеством изоморфных друг другу подграфов.

11) Чтобы данной проблемы не возникало, было принято решение оставить в графе только те подграфы, которые не изоморфны никаким другим подграфам этого графа, то есть подграфы с уникальной структурой.

12) Нужен ли вообще?

13) Для графа, состоящего из уникальных подграфов, обучающая выборка составлялась следующим образом: Для каждого подграфа считалась одна из представленных характеристик и на основе этой информации составлялась обучающая выборка.

Я отобрала только выделенные жирным характеристики, так как остальные характеристики связаны между собой и из-за них возникал эффект мультиколлиниарности.

14) Таких обучающих выборок три. Для каждой из обучающих выборок строится модель. Построенные модели показали следующие результаты. Количество регрессоров это количество посчитанных характеристик для уникальных графов после удаления нулевых и линейно зависимых столбцов. Количество значимых регрессоров – это кол-во регрессоров с уровнем значимости меньше 0.05

15) Чтобы улучшить модель я объединила эти три выборки,, исключила линейнозависимые столбцы и получила модель с коэфициентом детерминации 0.515

16) Также помимо обычного мнк я построила модель робастной мнк и для исследования возможности использования данной модели была посчитана кросс валидация**.** Результат показал, что робастная мнк имеет меньшую ошибку

17) Если сравнивать графики этих моделей, то субъективно можно заметить что робастный мнк гораздо лучше