

1. Аффективные_расстройства.

- Представляет наследственный фактор аффективных расстройств (редко встречаются) (депрессия, биполярное), значения - родственники с расстройством. Nan: 96.58% - можно увидеть слишком много пропусков, т.к данные болезни редки.
- Используем Ordinal encoding, потому что важна иерархия (важность).
- Выбор ordinal обоснован риском по родству (материнская линия усиливает генетический фактор). Примеры: 0 - пропуск/отсутствие (Nan); 1 - "Отец"/"Бабка по линии отца"; 2 - "Бабка по линии матери"/"Дед по линии матери"/"Дядя/Тётя по линии матери"; 3 - "Мать".

2. ПсихПоздВозр.

- Представляет наследственность психозов позднего возраста (деменция и т.д). Nan: 98.47%,- тоже слишком много пропусков, как и у Аффективных расстройств.
- Используем Ordinal encoding, подходит для весов по влиянию (поздние психозы наследуемы, материнская линия = выше).
- Примеры: 0-пропуск/отсутствие/"onehot"; 1 - "Дед по линии матери"/"Бабка по линии отца"; 2 — "Бабка по линии матери"/"мать".

3. НеврСтресс.

- Наследственность невротических/стрессовых расстройств (тревога, фобии). Nan: 99.17%.
- Ordinal encoding.
- Градация по родству (сиблинги/мать = высокий риск из-за генов); Примеры: 0 — пропуск; 1 — "Отец"/"Бабка по линии матери"/"Дядя/Тётя по линии матери"; 2 — "Мать"/"Сиблинги".

4. УО.

- Наследственность умственной отсталости. Nan: 99.88%.
- Label encoding (бинарный).
- Примеры: 0 — пропуск; 1 — "Сиблинги".

5. Эпилепсия.

- Наследственность эпилепсии. Nan: 98.94%.
- Ordinal encoding.
- Примеры: 0 - пропуск/"Отец"/"Дядя/Тетя по линии отца"/"Дед по линии матери"; 1 - "Дядя/Тётя по линии матери"; 2 - "Мать"/"Сиблинги".

6. ОргПорЦНС.

- Наследственность органических поражений ЦНС. Nan: 97.88%.
- Ordinal encoding с суммированием.
- В столбец разбиваем значения по запятой, чтобы для людей с несколькими проблемными родственниками значение было выше.

Примеры: 0 — пропуск/"Отец"/"Бабка по линии отца"; 1 — "Бабка по линии матери"/"Дядя/Тётя по линии матери"; 2 — "Мать"/"Сиблиниги"/"Дети" (если женщина; иначе 0).

7. Химическая зависимость.

- Наследственность химической зависимости (алкоголь/наркотики). Nan: 85.26%
- Ordinal encoding.
- Примеры: отсутствие, отец, дед по линии матери, дед по линии отца, бабка по линии отца - 0; бабка по линии матери, дядя/тётя по линии матери - 1; мать, сиблиниги - 2.

8. БерРодыМать.

- Представляет патологии беременности у матери. Nan: 4.95%, уникальных классов: 25, этот признак важен.
- Ordinal encoding.
- Градация по тяжести (токсикоз второй=тяжелее, невынашивание=макс); Примеры: 0 - "Без патологии"/"Не оценивались"/пропуск; 1 - "Токсикоз первой"; 2 - "Токсикоз второй"; 3 - "Инфекционные"/"Психотравмирующие"; 4 - "Патология"/"Другие тяжелые"; 5 - "Невынашивание".

9. ПатРодов.

Патологии родов. Nan: 5.42%, уникальных классов: 16

Ordinal encoding (объединить с ПатРодовДруг).

Объединение из-за дублирования (гипоксия=высокий риск ЦНС);

Примеры: 0 - "Нет"/пропуск; 1 - "Кесарево"/"Стремительные"; 2 -

"Затяжные"/"Щипцы"; 3 - "Преждевременные"; 4 -

"Гипоксия"/"Асфиксия"/"Эклампсия".

10. ПатРодовДруг.

Дополняет патологии родов (редкие уточнения). Nan: 96.23%.

Ordinal encoding с суммированием (объединить с ПатРодов).

Примеры: аналогично ПатРодов, 0 - "Нет"/пропуск; 1 -

"Кесарево"/"Стремительные"; 2 - "Затяжные"/"Щипцы"; 3 - "Преждевременные";

4 - "Гипоксия"/"Асфиксия"/"Эклампсия".

11. РанРазы.

Представляет раннее развитие (до 1 года). Nan: 4.01%, уникальных классов: 3.

Label encoding.

Шкала отклонения/нормы (отклонения=риск); Примеры: наличие отклонений - это -1; не оценивалось, пропуск в данных - 0; в соответствии с возрастом - 1.

12. Ут2.

Уточнение отклонений в раннем развитии. Nan: 95.17%.

Необычный столбец, так как в нем практически все значения уникальные, и более 90% - пропуски. В таком случае пропуски надо заменить, а для других изменений необходимо участие специалиста, который сможет оценить отклонения.

13. НевроРасстрДетство

Невротические расстройства детства. Nan: 80.54%, уникальных классов: 31
Ordinal encoding.

Примеры: 0 - пропуск; 1 - "Простые фобии"/"Эмоциональная"/"Парейдопические"; 2 - "Логоневроз"/"Энурез"/"Снохождение"; 3 - "Симбиотическая"/"Ипохондрические".

14. НевроРасстрШкола

Невротические расстройства школьного возраста. Nan: 4.48%, уникальных классов: 9.

Ordinal encoding.

Примеры: 0 — "В соответствии"/"Не оценивалось"/пропуск; 1 — "Гиперактивность"/"Дефицит"; 2 — "Отставание"/"Диссоциация"; 3 — "Трудности контактов"/"Наличие отклонений".

15. НевроРасстрШколаДоп

Уточнения к школьным расстройствам. Nan: 75.71%, уникальных классов: 42.
Ordinal encoding (объединить с НевроРасстрШкола).

Примеры: аналогично НевроРасстрШкола.

16. Симптомы_e3

Симптомы e3 (разорванность речи, статус/анамнез). Nan: 74.29%, уникальных классов: 8.

Ordinal encoding.

Примеры: 0 — пропуск/"Статус"/"Статус до лечения"; 1 — "Анамнез"; 2 — "Статус после лечения".

17. Симптомы_з2

Симптомы з2 (несообразность, статус/анамнез). Nan: 44.93%, уникальных классов: 9.

Ordinal encoding.

Аналогично e3 примеры: 0 — пропуск/"Статус"/"Статус до"; 1 — "Анамнез"; 2 — "Статус после".