

1. Аффективные_расстройства. Пропуски в столбце можно трактовать как отсутствие расстройства, то есть закодируем значением 0. Важно учитывать генетическую передачу. Чаще всего мать (18), затем отец (6). Поэтому предлагаем кодировку по степени влияния: отсутствие, отец, дед по линии матери, бабка по линии отца, дядя/тетя по линии отца - 0 (минимальное влияние); бабка по линии матери, дядя/тётя по линии матери - 1 (среднее); мать, сиблинги - 2.

2. ПсихПоздВозр. Пропусков очень много (808), расстройство редкое. Пропуски - 0. Уникальные упоминания минимальны (бабка по линии матери - 5 случаев). Кодируем: отсутствие или любые дальние родственники - 0; бабка по линии матери или аналогичные близкие по линии - 1.

3. НеврСтресс. Также крайне редкий фактор. Пропуски - 0. Упоминания: сиблинги и мать по 2 случая. Кодировка: отсутствие или отец - 0; по линии матери - 1; мать или сиблинги - 2.

4. УО. Только 1 упоминание (сиблинги). Столбец почти константный. Можно удалить, чтобы избежать переобучения или закодировать: пропуск - 0; сиблинги - 1.

5. Эпилепсия. Редко, но есть мать и сиблинги по 2 случая. Кодировка: отсутствие или отец/дальние - 0; дядя/тётя по линиям - 1; мать или сиблинги - 2.

6. ОргПорЦНС. Чаще отец (7), мать (4), бабка матери (4). Кодировка: отсутствие - 0; отец - 1; мать, сиблинги, бабка по материнской линии или дети - 2.

7. Хим. Зависимость. Самый частый фактор наследственности. Доминирует отец (72 случая), мать (11). Можно разбить по запятой. Кодировка: отсутствие или дальние (дед/бабка) - 0; дядя/тётя или бабка матери - 1; отец - 1 (чаще всего); мать или сиблинги - 2.

8. БерРодыМать Почти все пациенты имеют данные (только 8 пропусков). Пропуски - (0). Доминирует "без патологии". Разбить по запятой (множественные осложнения редки, но есть). Кодировка по частоте и типу: без патологии/не оценивалось - 0; токсикоз первой половины - 1; токсикоз второй - 2; патология беременности - 3; инфекции, невынашивание или тяжелые заболевания - 4. Суммировать при нескольких проблемах.

9. ПатРодов и ПатРодовДруг. В ПатРодов мало пропусков, в Друг - много. Пропуски/ "нет" - 0. Можно объединить в один признак (сумма). Разбить по запятой. Кодировка : нет - 0; кесарево/затяжные - 1; преждевременные - 2; гипоксия/асфиксия - 3; редкие тяжелые- 4. Суммировать.

10. Ранразы. Пропуски - 0. Кодировка: наличие отклонений -1; не оценивалось/пропуск - 0; в соответствии с возрастом - 1.

11. УТ2. Значения почти уникальные. Пропуски - 0. Для непустых: выделить ключевые слова (отставание/задержка - 1, судороги/гидроцефалия - 2, инфекции/травмы - 3). Суммировать.

12. НевроРасстрДетство. Чаще фобии потом привязанность и энурез. Кодировка: отсутствие - 0; простые фобии/логоневроз - 1; энурез/снохождение - 2; привязанность или эмоциональная лабильность - 3. Суммировать при множественности.

13. НевроРасстрШкола и НевроРасстрШколаДоп. Можем объединить в один. В Школа частые в соответствии ,отклонения. В Доп: трудности контактов и дефицит внимания. Кодировка: в соответствии/отсутствие - 0; отклонения без уточнений - 1; дефицит внимания/гиперактивность - 2; трудности контактов или отставание - 3.

14. Симптомы_е3. Пропусков 596. Значения: анамнез (204), статус (151), до/после лечения (редко). Кодировка: пропуск/статус до лечения - 0;

только анамнез или статус - 1; комбинации с до/после лечения - 2 (указывает на динамику)

15. Симптомы_з2. Пропусков меньше (347). Чаще анамнез и статус. Кодировка аналогично: пропуск/только статус - 0; анамнез - 1; статус до/после лечения или комбинации - 2 (больше данных о течении).