В данной работе стоит задача разработать алгоритм для предиктивного анализа степени тяжести дорожного происшествия. Задумываясь о реализации, в первую очередь были рассмотрены инструменты моделирования и прогнозирования как наиболее отвечающие потребностям. Среди них был выбран байесовский метод, так как классические статистические методы при высокой степени неопределённости не всегда дают достоверные результаты, и необходимо обращаться к более высокорезультативным методам прогнозирования, одним из которых является байесовский метод, основанный на экспертных оценках и априорных данных наблюдений. Это вероятностная модель, представляющая собой множество переменных и их вероятностных зависимостей по Байесу. Байесовская сеть же способна как предсказывать результирующие показатели, так и возможные альтернативы для совокупности взаимодействия некоторых факторов. Поэтому байесовская сеть для данного случая подходит больше других методов, т.к. используется для прогнозов в условиях неопределённости

Рассмотрим подробнее формулу Байеса на примере в рамках данной работы. Пусть P(A) – априорная вероятность гипотезы A. Априорной вероятностью называется безусловная вероятность наступления гипотезы. P(B) – полная вероятность наступления события B., P(A|B) – вероятность гипотезы A при срабатывании события B (апостериорная вероятность), P(B|A) – вероятность наступления события B в случае истинности гипотезы A. В своём классическом виде формула Байеса имеет следующий вид:

Однако нас в данной работе интересует другой вид данной формулы, т.к. вероятность P(B) в задачах вычисляется как правило по формуле полной вероятности события, зависящего от нескольких несовместных гипотез, имеющих суммарную вероятность 1. В таком случае формула Байеса принимает вид:

Таким образом, используя формулу Байеса можно прогнозировать не только вероятность наступления результирующего события или гипотезы, но и напротив, стартовых событий, гипотез, из-за чего формула Байеса эффективна в условиях неопределённости.

Рассмотрим использование формулы Байеса на примере нашего исследования: в один месяц случилось 50 аварий с легковыми автомобилями, 30 аварий с грузовиками и 20 аварий с автобусами, во второй месяц случилось 30 аварий с легковыми автомобилями, 10 аварий с грузовиками и 30 аварий с автобусами. Данные по месяцам лежат в одной базе данных не рассортированными. Мы берём наугад один из случаев, в нём авария произошла с легковым автомобилем. Какова вероятность того, что данная авария произошла в первом месяце? Всего 170 аварий. Событие B – искомая машина – легковая. Гипотеза Aj – случай был взят с j-го месяца собираемой нами базы данных ДТП. Тогда:

Вероятность события B равна:

Данный пример наглядно иллюстрирует применение Байесовского метода, однако также он иллюстрирует её недостаток, при котором мы в данном случае учитываем лишь одно условие, в то время как в реальной жизни на различные события и гипотезы воздействие указывает целая совокупность различных условий, зачастую зависимых друг от друга. В таком случае применяется байесовская сеть, всё сводится к той же формуле байеса, но за счёт расчёта влияния всех условий на конечный прогноз, тот становится куда более близким к действительности