

К. А. Осипенко, А. Н. Осипенко, Н. Б. Осипенко
(ГГУ им. Ф. Скорины, ГГТУ им. П.О. Сухого, ИПК и ПК, Гомель)

РАСПОЗНАВАНИЕ ГРУПП РИСКА СМЕРТИ НА ОСНОВЕ ПАСПОРТНЫХ ДАННЫХ

В настоящей работе предлагается разработать методологию и программно-технологическое обеспечение предварительной экспресс-диагностики на основе паспортных сведений, данных о группе крови, антропометрии, анамнезе и иной стандартной информации, имеющейся в поликлинических базах данных..

В результате такой диагностики для каждого человека будут сформированы оценки вероятностей принадлежности к основным группам риска. В качестве исходного материала для пробного исследования возможностей использования паспортных сведений на первом этапе скрининга послужила выборка из 88 знаменитых людей. По каждому человеку из интернета брались данные в виде строки: имя; фамилия; страна, в которой жил человек; тип менталитета; день, месяц и год рождения; день, месяц и год смерти; количество жен (мужей); количество детей (включая приемных); основной род занятий по жизни (профессия); общая причина смерти (заболевание сердечнососудистой системы – инфаркт или инсульт – 1, онкозаболевание – 2, другие болезни – 3, несчастные случаи – 4, преднамеренное убийство – 5, иное – 6).

Для обработки данных использовался пакет «Statistica». В рамках этого пакета была написана программа обучения распознаванию групп риска. Перевод даты рождения в набор числовых качественных признаков осуществлен с помощью общеизвестного алгоритма Пифагора и описан в работе [1]. Для перевода имени и фамилии в признаки для распознавания групп риска использована числовая азбука [2].

Обучение алгоритма распознавания осуществлялось для двух основных групп риска. В первую группу вошли умершие от сердечнососудистых заболеваний и от несчастных случаев. Во вторую – умершие от онкозаболеваний и других хронических болезней.

Оценка качества полученной в результате обучения дискриминантной функции проводилась по исходной выборке. При этом ошибка первого рода (отнесение объекта первого класса ко второму) составила 10/55, ошибка второго рода – 1/33. Доля отказов – 39/88.

Высокую ошибку распознавания первой группы (в основном это смерти от сердечнососудистой болезни) можно объяснить неполнотой диагноза смертельного исхода, особенно в зрелом возрасте, когда инфаркт или инсульт становится следствием какой-либо хронической болезни. Часто это не афишируется даже для знаменитостей и никак не отражается в диагнозах смерти.

Несмотря на небольшой объем выборки в проведенном пробном исследовании, можно утверждать, что паспортные данные вполне пригодны для включения их в список признаков при экспресс-диагностике группы риска на первом этапе скрининга здоровья населения.

Литература

1 Осипенко, К.А. Метод регрессионного моделирования продолжительности жизни по дате рождения / К.А.Осипенко, Н.Б.Осипенко // Творчество молодых 2012: сборник научных работ студентов и аспирантов УО «ГГУ им. Ф. Скорины»: в 2 ч. / Гомельский гос. ун-т им. Ф.Скорины; отв. ред. О.М. Демиденко. – Гомель, 2012. –Ч. 1. –С.194-197.

2 Хигир, Б.Ю. Число имени / Б.Ю. Хигир. –СПб.: Астрель, 2008. –42с.