Министерство науки и высшего образования Российской Федерации
Федеральное государственное бюджетное образовательное учреждение высшего образования
«Московский государственный технический университет имени Н.Э. Баумана
(национальный исследовательский университет)»
(МГТУ им. Н.Э. Баумана)

ФАКУЛЬТЕТ «Информатика и системы управления»

КАФЕДРА «Программное обеспечение ЭВМ и информационные технологии»

РАСЧЕТНО-ПОЯСНИТЕЛЬНАЯ ЗАПИСКА К НАУЧНО-ИССЛЕДОВАТЕЛЬСКОЙ РАБОТЕ НА ТЕМУ:

"Классификация известных методов противодействия мошенничеству в области больших данных"

Студент Коняев Е.А

Группа ИУ7-53Б

Научный руководитель Гаврилова Ю.М.

Подпись руководителя_____

Москва — 2022 г.

Оглавление

Введение

Существуют различные типы рисков в финансовой сфере, такие как финансирование терроризма, отмывание денег, мошенничество с кредитными картами и мошенничество со страховкой, которые могут привести к катастрофическим последствиям для таких организаций, как банки или страховые компании. Причем экономические последствия мошенничества могут быть более серьезными, чем просто материальные затраты. Не менее значимыми показателями убытков от мошенничества являются потеря репутации, потеря доверия потребителей, потеря налогового дохода и т.д. Именно поэтому крайне важно противостоять мошенничеству в финансовой сфере.

Вышеописанные финансовые риски обычно выявляются с помощью алгоритмов классификации. В задачах классификации асимметричное распределение классов, также известное как дисбаланс классов, является очень распространенной проблемой при обнаружении финансового мошенничества, когда для решения этой проблемы используются специальные подходы к анализу данных наряду с традиционными алгоритмами классификации. Эта проблема становится более значительной, когда мы рассматриваем область Big Data, методы которой применяются для анализа баз данных в финансовой сфере, так как обычноми методами обработки реляционных таблиц это уже невозможно сделать в силу масштабности пользовательских данных. Поэтому важно при выявлении мошеннических действий правильно вибрать методы анализа из области BigData, которые будут эффективны для данной задачи.

С учетом всего вышесказанного, целью данной является анализ известных методов противодействию мошенничества в области BigData. Для этого необходимо:

1) провести анализ предметной области;

2) сформулировать критерии оценки известных методов методов;

3) провести обзор методов;

4) классифицировать методы.

Глава 1

Аналитическая часть

1.1 Анализ предметной области

1.2 Общая характеритика Big Data

Термин «Big Data» означает большие работы (коллекции, потоки) данных, которые не могут быть обработаны традиционными компьютерными техниками. Этот термин означает не само понятие «большие данные», а предмет исследования, который включает в себя различные инструменты, техники и платформы.

Большие данные включают в себя информацию, генерируемую различными системами и приложениями. Некоторые из сфер, которые попадают под определение «Big Data»:

1) черный ящик: информационная составляющая часть вертолета, самолета, морского/-космического корабля. Данные подобного рода включают в себя запись голосов экипажа (микрофоны и наушники), информацию о характеристиках объекта управления;

2) социальные медиа: включают данные, распространяемые через социальные сети;

3) фондовые биржи: хранение информации о сделках купли-продажи между копаниями-партнерами;

4) транспортные системы: модели, характеристики, расстояния - все информация о транспорте и дорожных сетях.

5) и т.п.

Как следствие, термин «Big Data» включает большое объем, высокую скорость обработки и широкое разнообразие данных и делится на три типа:

1) структурные данные – реляционные БД;

2) полу-структурированные данные – XML-файлы;

3) неструктурированные данные – файлы формата Word, PDF, Text, медиа-журналы.

Для использования возможностей больших данных требуется инфраструктура, которая может управлять и обрабатывать огромные объемы структурированных и неструктурированных данных в реальном времени. Существуют два класса техники, которые обрабатывают

большие данные: технологии обработки Big Data и программно-аппаратные средства работы с большими данными.

Технологии обработки Big Data:

1) модель распределенных вычислений MapReduce;

2) технологии Hadoop;

3) подход NoSQL;

4) язык программирования R.

Программно-апаратные средства работы с Big Data:

1) комплекс инструментов Oracle Exalytics;

2) аппаратно-программный комплекс SAP HANA;

3) IBM Watson Explorer.

1.3 Обнаружение мошенничества в области Big Data

Аналитика мошенничества на основе больших данных является эффективным подходом по трем причинам. Эти причины сводятся к следующему:

1) **точность**. Большинство организаций имеют ограниченные возможности для проверки случаев мошенничества человеком. Цель аналитики мошенничества на основе данных состоит в том, чтобы наиболее оптимально использовать имеющиеся ограниченные возможности или в другими словами, чтобы максимизировать долю мошеннических дел среди проверенных (и возможно обнаруженную сумму мошенничества);

2) **операционная эффективность**. Аналитика мошенничества на основе данных построена с использованием методов из различных областей, включая машинное обучение, статистику, математику, глубокое обучение. Это значительно повышает эффективность работы что критически важно для многих сценариев мошенничества. Например, при оценке транзакция с кредитной картой, требуется почти немедленное решение относительно одобрить или заблокировать транзакцию из-за подозрения в мошенничестве;

3) **экономичность**. Разработка и поддержка эффективной и экономичной системы обнаружения мошенничества, основанной на человеческих русурсах, является сложной и трудоемкой задачей. Более автоматизированный и более эффективный подход к разработке и обслуживанию системы обнаружения мошенничества является подход, основанный на анализе больших данных.

Подводя итог вышесказанному, становится очевидно, что аналитика мошенничества на основе данных имеет много преимуществ по сравнению с традиционным экспертным подходом к обнаружению мошенничества. Однако в этом методе обнаружения мошенничества существуют различные проблемы, которые необходимо решить или минимизировать, чтобы усилить эффективность такой системы. Одной из таких критических проблем является проблема дисбаланса классов.