Слайд 1

Здравствуйте, уважаемые члена комиссии. Меня зовут Султанов Азамат.

Тема моего исследования - Обнаружение внутреннего нарушителя путём выявления стрессового состояния пользователя

Слайд 2

В ходе анализа предметной области было найдено несколько работ, в которых авторы показали, что выявление стресса с использованием различных биометрических показателей позволяет обнаруживать внутреннего нарушителя. В данных работах биометрические показатели отслеживались с помощью специального дорогостоящего оборудования, которое зачастую очень сложно найти, например, в обычном офисе или дома, что является причиной усложнённой интеграции исследованных методов на практике.

Однако есть метод, который интересен именно с точки зрения перспективы интеграции, в силу доступности датчиков, в роли которых выступают клавиатура и мышь. К сожалению, именно в данной области найдена всего лишь одна работа, где в качестве биометрических показателей используется динамика взаимодействия с клавиатурой и мышью. Именно это побудило меня на развитие и усовершенствование идеи использования клавиатуры и мыши для детекции стресса, поэтому целью моей работы стало оценить возможность выявления стрессового состояния пользователя на основе анализа взаимодействия с клавиатурой и мышью.

Слайд 3

В силу отсутствия датасетов в исследуемой области, возникла необходимость организации процесса сбора данных.

Для этого были предложены сценарии, которые были спроектированы таким образом, чтобы максимально близко описать ситуации, которые могли бы произойти в реальности. Часть этих сценариев описывает случаи правомерного поведения сотрудника, которые не сопровождаются стрессом, другая описывает ситуации, в которых имитируются действия внутреннего нарушителя, которые предполагают индукцию стресса, в силу наложенных временных ограничений.

Для сбора данных было написано специальное ПО, фиксирующее в фоновом режиме события мыши и клавиатуры в момент выполнения сценариев участниками эксперимента.

Слайд 4

Так как анализ данных в работе основан на применении алгоритмов машинного обучения, то дальнейшим шагом стало выделение признаков из файлов логирования. Были вычислены, как временные, так и частотные признаки.

В результате выделения признаков был получен датасет на основе которого производилось дальнейшее исследование.

Слайд 5

Перед началом обучения моделей были проделаны простые шаги предобработки данных, включающие удаление признаков, связанных с редкими события и заполнение пустот в датасете медианами соответствующих признаков.

Слайд 6

На текущих графиках приведены усреднённые значения признаков, которые были по отдельности рассчитаны для категорий нормального и аномального поведений.

Красными зонами выделены наиболее информативные признаки. Они считаются наиболее информативными, потому что именно в этих зонах наблюдается существенное различие признаков, описывающих нормальное и аномальное поведение.

В дальнейшем планируется применить алгоритмы выделения наиболее информативных признаков.

Слайд 7

С использованием итогового набора признаков были обучены модели бинарных классификаторов на основе различных алгоритмов машинного обучения. Для сравнения информативности, признаки были разделены на различные наборы. Сначала модели обучались только на признаках, выделенных для мыши. Далее были добавлены признаки специальных клавиш, далее обучение происходило на всём наборе признаков.

В первом случае модели недообучились. Причиной этому стало то, что мышь использовалась крайне редко при выполнении сценариев, поэтому события мыши не обладают сильной описательной способностью поведения. Во втором случае получены наилучшие результаты, и это объясняется тем, что были добавлены события клавиатуры.

В последнем случае точность предсказаний упала по сравнению со вторым случаем, так как количество признаков сильно увеличилось, но количество данных осталось прежним и тут надо применять алгоритмы выделения информативных признаков.

Слайд 8

В ходе проведённого исследования были:

* Проанализированы существующие методы обнаружения внутреннего нарушителя с использованием биометрических показателей на основе алгоритмов машинного обучения
* Предложены сценарии поведения и собраны данные
* Реализованы процессы накопления данных, предобработки данных, обучения и оценки моделей классификаторов
* Наилучшие результаты по выявлению стресса получены для моделей на основе алгоритмов логистической регрессии и метода k-ближайших соседей