

<u>234053</u>
<i>Numer indeksu</i>
<u>Paweł Galewicz</u>
<i>Imię i nazwisko</i>

<u>234067</u>
<i>Numer indeksu</i>
<u>Bartosz Jurczewski</u>
<i>Imię i nazwisko</i>

<u>234102</u>
<i>Numer indeksu</i>
<u>Zbigniew Nowacki</u>
<i>Imię i nazwisko</i>

<u>234106</u>
<i>Numer indeksu</i>
<u>Karol Podlewski</u>
<i>Imię i nazwisko</i>

<u>234128</u>
<i>Numer indeksu</i>
<u>Piotr Wardecki</u>
<i>Imię i nazwisko</i>

Kierunek	Informatyka Stosowana
Stopień	II
Specjalizacja	Data Science
Semestr	1

Data oddania 29 kwietnia 2020

Metody uczenia maszynowego

Analiza danych przestępczych

Propozycja projektu grupowego

1 Tytuł

Projekt jaki zamierzamy wykonać został zatytułowany **Analiza danych przestępczych**.

2 Dane członków grupy

Skład grupy:

- Paweł Galewicz 234053
- Bartosz Jurczewski 234067
- Zbigniew Nowacki 234102
- Karol Podlewski 234106
- Piotr Wardęcki 234128

3 Motywacja

Klasyfikacja danych pozwala połączyć przypadki, które dla ludzkiego oka nie wydają się mieć ze sobą wiele wspólnego. Taka analiza może się okazać niezwykle przydatna nie tylko w przypadku danych medycznych, ale też przestępczych. Wiedząc co łączy nietypowe rabunki czy akty wandalizmu możemy lepiej zaplanować działania prewencyjne, a także lepiej chronić nasze mienie jak i samych siebie.

4 Cele projektu

Celem projektu jest przeprowadzenie badania mającego na celu zwiększenie efektywności działań prewencyjnych służb mundurowych. Program zostanie wykonany przy użyciu języka programowania Python wraz z różnymi dostępnymi frameworkami pozwalającymi na szybką i skuteczną implementację niezbędnych algorytmów. Wykonanie projektu pozwoli nam na lepsze zrozumienie działania algorytmów oraz wzrost umiejętności pracy z dużymi zbiorami danych które są wykorzystywane w rzeczywistych warunkach. Dodatkowymi korzyściami będzie nauka pisania czytelniejszego kodu w składni języka Python, a także lepsze poznanie wielu bibliotek pomocniczych, tylko pośrednio związanych z analizą danych. Wykonanie projektu powinno się zakończyć wyciągnięciem i przedstawieniem uzyskanych wniosków na początku czerwca.

5 Opis zbiorów danych

Dane zostaną pozyskane z wykorzystaniem najpopularniejszych repozytoriów danych (m.in. Kaggle, UCI Repository czy Makeover Monday). Zostanie podjęta próba odszukania danych na stronach anglo- i polskojęzycznych, które są bezpośrednio lub pośrednio powiązane z publicznymi jednostkami monitorującymi przestępstwa (na przykład strony rządowe czy strony funkcjonariatów policji). Scenariusz optymistyczny zakłada pozyskanie danych zbieranych na przestrzeni wielu lat. Takie dane potencjalnie pomóc mogą w odkryciu interesujących trendów, które pozwolą na głębsze zrozumienie zjawiska przestępczości oraz gdzie może ono zmierzać w przyszłości.

Dodatkowo planowane jest zestawienie danych o występach ze statystykami miejsc, w których dane zjawiska występują. Poszerzy to pogląd na dane i pozwoli na poszerzenie poglądu na informacje o otoczeniu i społeczności, w której występkę mają miejsce.

6 Plan czynności

1. Pozyskanie jak największej liczby danych o przestępczości
2. Zapoznanie się ze zgromadzonym zestawem i ocena jego użyteczności
3. Czyszczenie zbioru danych oraz próba wyekstrahowania najistotniejszych cech
4. Zestawienie danych o przestępczości z danymi statystycznymi miejsc ich występowania
5. Analiza pod kątem niestandardowych, obiecujących korelacji danych, które warto będzie dalszego eksplorowania
6. Analiza skupień w celu znalezienia wzorców
7. Analiza częstotliwości oraz wystąpień w czasie, określenie tendencji i próba predykcji przyszłych wartości
8. Porównanie obszarów z największą częstością występowania zjawisk przestępczości w czasie