**SSIED Projekt**

Pump it Up

Martin Czakański

Adam Krawiec

Kuba Łada

Michał Slabik

## Tomasz Ścigała

Wykonane Kroki:

1. Rejestracja oraz utworzenie zespołu w konkursie: [Pump it Up](https://www.drivendata.org/competitions/7/pump-it-up-data-mining-the-water-table/team/)
2. Utworzenie repozytorium [GitHub](https://github.com/martinCzakanski/SSIED-DrivenData-PumpItUp)
3. Pobranie danych do analizy
4. Wstępna analiza danych
5. Wybranie technologi w której bedzię realizowany projekt - RapidMiner
6. Utworzenie procesu w RapidMiner
7. Opis zawartości plików źródłowych

Dalsze kroki:

1. Porównanie zakresu cech między Danymi treningowymi a testowymi
2. Oczyszczenie danych (wartości domyślne, duplikacje itp)
3. Analiza cech pod kątem konwersji typu danych
4. Wyznaczenie kluczowych cech
5. TODO

# **Wprowadzenie**

//TODO – Opis celu projektu

# **Opis zawartości plików źródłowych**

# SubmissionFormat.csv oraz Training set values.csv

* + - Id – Id pompy wodnej
    - status\_group - Stan pompy wodnej możliwe wartości:
      * functional
      * non functional
      * functional needs repair

Test set values.csv oraz Training set values.csv

* Id -
* amount\_tsh -
* date\_recorded -
* funder -
* gps\_height -
* installer -
* longitude -
* latitude -
* wpt\_name -
* num\_private -
* basin -
* subvillage -
* region -
* region\_code -
* district\_code -
* lga -
* ward -
* population -
* public\_meeting -
* recorded\_by -
* scheme\_management -
* scheme\_name -
* permit -
* construction\_year -
* extraction\_type -
* extraction\_type\_group -
* extraction\_type\_class -
* management -
* management\_group -
* payment -
* payment\_type -
* water\_quality -
* quality\_group -
* quantity -
* quantity\_group -
* source -
* source\_type -
* source\_class -
* waterpoint\_type -
* waterpoint\_type\_group -