

# Projekt Transformatory

## CEL

Celem projektu jest wskazanie efektywnej polityki konserwacji transformatorów średniego napięcia przy zastosowaniu metodyki nauki o danych.

## DANE

- Zbiór danych *awarie\_transf\_hist.xlsx*
- Zbiór danych *awarie\_transf\_new.xlsx*

## PROBLEM BIZNESOWY

Awaria sprzętu jest nieuniknioną rzeczywistością wszystkich przedsiębiorstw świadczących usługi z wykorzystaniem infrastruktury technicznej. Konserwacja zapobiegawcza jest najlepszym sposobem na uniknięcie awarii i związanych z nimi opóźnień w obsłudze klienta, kosztownych napraw, kar umownych. Taka konserwacja może być przeprowadzana regularnie w przypadku wszystkich urządzeń, ale może to nie być rozwiązanie efektywne kosztowo. Konserwacja "predykcyjna" określa, które urządzenia mogą ulec awarii przy danym zestawie ograniczeń (takich jak okres czasu, zdarzenie i obciążenie).

### Opis otoczenia biznesowego

Przedsiębiorstwo E (operator sieci elektroenergetycznej) dysponuje zestawem danych historycznych dotyczących awarii transformatorów w stanie Georgia w określonym czasie (por. opis danych). Wiadomo, że konserwacja zapobiegawcza jednego transformatora kosztuje 30 tys. USD. Dodatkowo wiadomo, że w przypadku zaistnienia awarii przeciętny koszt wynikający z utraconych przychodów, kar umownych oraz naprawy awaryjnej transformatora wynosi 100 tys. USD. Dysponując danymi historycznymi oraz danymi dotyczącymi zbioru urządzeń, które nie były objęte monitoringiem, przedsiębiorstwo chce zrealizować następujące cele biznesowe:

Cele biznesowe:

- wskazanie najważniejszych czynników zwiększających prawdopodobieństwo zaistnienia awarii transformatora,
- ustalenie progu dotarcia (odsetka populacji transformatorów), przy którym wykorzystanie predykcji awarii jest opłacalne (lepsze niż objęcie konserwacją zapobiegawczą całej populacji),
- określenie optymalnej ze względu na koszty polityki konserwacji dla 20 kolejnych urządzeń w stanie Georgia (dane *awarie\_transf\_new.xls*).

## Opis danych

cecha	znaczenie
<i>Status</i>	czy transformator uległa awarii (target)
<i>PMLate</i>	czy konserwacja była opóźniona w relacji do ресурсu urządzenia
<i>Overloads</i>	liczba przeciążeń urządzenia
<i>MilesFromOcean</i>	odległość od linii brzegowej
<i>Manufacturer</i>	producent
<i>WaterExposure</i>	czy urządzenie jest narażone na działanie wody
<i>MultipleConnects</i>	czy urządzenie jest hubem
<i>Storm</i>	czy urządzenie narażone było na gwałtowne zjawisko atmosferyczne
<i>AssetType</i>	typ urządzenia
<i>Repairs</i>	naprawy (czy i jakie)
<i>AssetId</i>	identyfikator urządzenia
<i>AssetLocation</i>	lokalizacja urządzenia - ulica
<i>AssetCity</i>	lokalizacja urządzenia - miasto
<i>AssetState</i>	lokalizacja urządzenia - stan
<i>AssetZip</i>	kod pocztowy lokalizacji
<i>Lat</i>	szerokość geograficzna
<i>Long</i>	długość geograficzna
<i>AvgRepairCost</i>	przeciętny koszt naprawy po awarii
<i>Age</i>	wiek urządzenia

## NARZĘDZIA

W zakresie EDA (wstępne badanie danych), przygotowania danych, wizualizacji – dowolne poznane narzędzie.

W zakresie budowy, szacowania i aplikacji właściwego mechanizmu uczącego się – sugerowane narzędzie SAP Predictive Analytics Automated Mode lub inne poznane.

## EFEKT PRACY NAD PROJEKTEM

Efektem pracy nad projektem będzie raport z wykonania projektu.

W projekcie należy wykorzystać wiedzę z zakresu organizacji procesu eksploracji danych (schemat procesu CRISP-DM). Działania wykonane dla etapów należy dokumentować. Do raportu należy załączyć:

- opracowany zbiór (zbiory) danych użyte na etapie modelowania,
- kody źródłowe (np. notebooki Jupyter) lub zapisane skrypty wykonywalne SAP PA, umożliwiające replikację wyników.

Format pliku głównego raportu – .pdf. Całość raportu tj. plik główny oraz załączniki należy scalić w archiwum .zip.