

Hadoop & Spark

dr Mariusz Rafało

<http://mariuszrafalo.pl>

mrafalo@sgh.waw.pl

Zasady współpracy dla prac dyplomowych (studia podyplomowe)

- Proszę zapoznać się dokumentem a następnie przesłać następujące informacje dotyczące planowanej pracy:
 - Temat (jeśli znany) lub ogólna tematyka pracy
 - Poziom znajomości technologii IT (programowanie, chmura, analiza danych). **Uwaga:** nie jest wymagane aby posiadać kompetencje IT, ale potrzebuję wiedzieć aby zaplanować zakres pracy
- Jeśli będzie potrzeba spotykamy się celem skonsultowania i wypracowania tematu.
- Przed rozpoczęciem pisania ważne jest abyśmy mieli potwierdzony ramowy plan pracy: rozdziały i podrozdziały.
- W trakcie pracy nad dokumentem komunikujemy się drogą mailową lub spotykamy się na Teams jeśli jest potrzeba konsultacji.

Tematyka prac

Prowadzę prace dyplome obejmujące następujące zagadnienia:

1. Przetwarzanie danych przy wykorzystaniu technologii big data
2. Zastosowanie big data w różnych dziedzinach (medycyna, przemysł, marketing, itp.)

3. Przetwarzanie i analizowanie danych w czasie rzeczywistym
4. Hurtownie danych i Business Intelligence
5. AI i uczenie maszynowe

Przykładowe wypromowane prace

1. Wpływ zjawisk pogodowych na rozwój szkód ubezpieczeniowych w przyszłości
2. Możliwości wykorzystania big data w działalności najwyższych organów kontroli
3. Analiza sentymentu na przykładzie wpisów polskich polityków na platformie Twitter w porównaniu z wynikami sondaży wyborczych
4. Przetwarzanie oraz analiza danych w czasie rzeczywistym na podstawie API ZTM Warszawa
5. Analiza wyników meczów tenisa kobiet w latach 2013-2022 przy wykorzystaniu języka Python i środowiska Databricks
6. Big Data w centrum logistycznym – analiza danych sprzedażowych i wysyłkowych w przedsiębiorstwie z branży AGD i RTV
7. Porównanie skuteczności modeli predykcyjnych w wykrywaniu klientów, którzy nie wywiążą się ze spłaty zobowiązania
8. Analiza danych w badaniach klinicznych z przykładem analizy w języku Python
9. Analiza danych demograficznych z wykorzystaniem Apache Spark

Wytyczne edycyjne

Struktura pracy

Praca zazwyczaj składa się ze wstępu (nienumerowanego), trzech rozdziałów oraz zakończenia (nienumerowanego)

- We wstępie piszemy o czym jest praca, jaki jest cel pracy, struktura pracy
- Rozdział 1 - teoretyczny
- Rozdział 2 - teoretyczny
- Rozdział 3 - praktyczny, opisujący analizę danych, zbudowany system, itp. (o ile praca ma charakter praktyczny)
- W zakończeniu piszemy: wnioski co wyszło, ograniczenia pracy, kierunek dalszych badań, możliwości zastosowania praktycznego wyników pracy

Edycja dokumentu

1. Praca powinna mieć około 45 stron
2. Praca powinna mieć 3 rozdziały oraz wstęp i zakończenie. Rozdziały podzielone na podrozdziały
3. Praca powinna zawierać odniesienia do literatury w treści (szczególnie ważne są w rozdziałach teoretycznych)
4. Na końcu pracy bibliografia; powinna mieć ok 30 pozycji
5. Dobrze jakby rozdziały miały porównywalną długość, ok. 13-15 stron
6. Każdy rysunek/tabela musi mieć numer, tytuł, źródło oraz odniesienie w treści (po numerze)
7. Tabele proszę umieszczać w formie tabel a nie obrazków