

AI Slut project ITH-Distans

Machine learning genom Energi mättning inom Industri

Contents

Rapport: Förberedande Analys och Förståelse av Data för Energiadministration i Industriella	
Sammanhang	3
Inledning:	3
Dataförståelse:	3
Datakompletthet:	3
Null-värden:.....	3
Extrema Värden:.....	3
Datatyper:	3
Relevanta Datafält:	3
Konvertering till Numeriskt Format:	3
Förberedande Åtgärder:.....	3
Slutsats:.....	4
Rapport: Genomgång av Dataförberedelseprocess för Energiadministration i	
Industriella Sammanhang	5
1. Dataförståelse:	5
2. Datakompletthet:	5
3. Hantering av Null-värden:	5
4. Identifiering och Hantering av Extrema Värden:	5
5. Analys av Datatyper:	5
6. Urval av Relevanta Datafält:	5
7. Konvertering till Numeriskt Format:	5
8. Explorativ Dataanalys (EDA):	6
Slutsats och Nästa Steg:	6
Rapport: Verktyg och Kod Dokumentation för Dataanalys i Energiadministration.....	7

Inledning:	7
Verktyg som Används i Projektet:	7
Datainsamling och Lagring:	7
Dataförädling och Analys:	7
Versionskontroll och Samarbete:	7
Dokumentation av Kod för Prediktivt Underhåll Baseras på Drivdata	8
Översikt:	8
Kodbibliotek:	8
Importera Data:	8
Datarengöring:	8
Introducera Anomalier:	8
Funktionstillverkning:	8
Normalisering av Data:	8
Förbereda Tränings- och Testdata:	8
Modellträning:	9
Modellutvärdering:	9
Funktionens Viktighet:	9
Spara Modellen (Valfritt):	9
Användningsinstruktioner:	9

Rapport: Förberedande Analys och Förståelse av Data för Energiadministration i Industriella Sammanhang

Inledning:

Denna rapport syftar till att etablera en strukturerad process för att förstå och förbereda min dataset inför en fördjupad analys med syfte att förbättra energihanteringen inom industrin. Innan jag utforskar specifika tillämpningar av maskininlärning är det avgörande att jag noggrant granskar och förstår min data.

Dataförståelse:

En grundlig dataförståelse innebär att jag undersöker följande aspekter:

Datakompletthet:

Är min dataset heltäckande för den tidsperiod och de parametrar jag avser analysera? Om data saknas, hur påverkar det vår analys?

Null-värden:

Finns det null-värden i min dataset? Om så är fallet, behöver jag fastställa strategier för att hantera dessa, exempelvis genom interpolering eller borttagning av ofullständiga dataposter.

Extrema Värden:

Behöver jag identifiera och utvärdera extrema värden (outliers) i datat? Extrema värden kan antingen vara indikatorer på datafel eller jagtliga insikter om ovanliga händelser.

Datatyper:

Jagliga datatyper representerar min data? Förståelse för olika datatyper (numeriska, kategoriska, tidsmässiga etc.) är avgörande för korrekt datahantering och analys.

Relevanta Datafält:

Jagliga specifika fält i min dataset är relevanta för vår energihanteringsanalys? Jag behöver identifiera och prioritera dessa för att fokusera vår analys.

Konvertering till Numeriskt Format:

Hur kan jag omvandla icke-numeriska datafält till ett format som kan hanteras av våra analytiska modeller? Metoder för detta kan inkludera kodning av kategoriska variabler eller normalisering av data.

Förberedande Åtgärder:

För att säkerställa att min dataset är lämpligt för avancerad analys, bör följande steg övervägas:

Rengöring och förbehandling av data för att hantera null-värden och outliers.

Konvertering av relevanta data till en enhetlig och analytiskt användbar form.

Utförande av en explorativ dataanalys (EDA) för att få en djupare förståelse av datamönster och -strukturen.

Slutsats:

Denna inledande fas av dataförståelse och förberedelse är kritisk för framgången av vår efterföljande analys. Genom att noggrant granska och förbereda min data, lägger jag grunden för en effektiv och insiktsfull användning av maskininlärning för att optimera energihanteringen i vår industriella verksamhet.

Rapport: Genomgång av Dataförberedelseprocess för Energiadministration i Industriella Sammanhang

Sammanfattning:

Denna rapport beskriver den genomgånga processen för förberedelse och analys av min dataset, med målet att optimera energihantering inom industriell verksamhet. Processen innefattar steg för dataförståelse, rengöring, och förbehandling inför tillämpning av avancerade analytiska metoder inklusive maskininlärning.

1. Dataförståelse:

Insamling av Data: Jag började med att samla in data från olika källor inom vår industriella miljö, inklusive energiförbrukningsdata, produktionsloggar och miljösensorer.

Initial Granskning: Ett första steg var att granska datasettet för att få en överblick över dess struktur, omfattning och kvalitet.

2. Datakompletthet:

Jag utförde en genomgående kontroll för att identifiera eventuella brister i datat. Detta inkluderade sökning efter saknade värden över tidsperioder och maskiner.

3. Hantering av Null-värden:

Strategier utvecklades för att hantera null-värden, där jag antingen ersatte dessa med genomsnittsvärden, medianvärden eller genom interpolation, beroende på kontexten.

4. Identifiering och Hantering av Extrema Värden:

Extrema värden undersöktes för att fastställa om de var resultat av datafel eller verkliga händelser. När det var nödvändigt, korrigerades eller uteslöts dessa värden.

5. Analys av Datatyper:

Varje datatyp i min dataset granskades för att säkerställa korrekt hantering och analys. Detta inkluderade omvandling av tidsstämplar och kategorisering av textdata.

6. Urval av Relevanta Datafält:

Jag valde ut de mest relevanta datafälten baserat på deras potential att bidra till insikter kring energiförbrukning och effektivitet.

7. Konvertering till Numeriskt Format:

Icke-numeriska data omvandlades till numeriska format genom tekniker som one-hot encoding för kategoriska variabler och skalning för proportionella data.

8. Explorativ Dataanalys (EDA):

Ett utförligt EDA utfördes för att upptäcka mönster, korrelationer och insikter i datat, vilket gav en djupare förståelse för hur olika faktorer påverkar energiförbrukningen.

Slutsats och Nästa Steg:

Genom denna noggranna förberedelse och förståelse av min data, har jag lagt en robust grund för den efterföljande fasen där jag kommer att implementera och utvärdera maskininlärningsmodeller. Denna process kommer att möjliggöra mer precisa förutsägelser och optimeringar gällande energianvändning och effektivitet i våra industriella operationer.

Rapport: Verktyg och Kod Dokumentation för Dataanalys i Energiadministration

Inledning:

Denna rapport ger en översikt över de verktyg som använts och dokumentationen av kod som utarbetats under vårt projekt för att förbättra energihanteringen i en industriell miljö. Projektet innebar omfattande dataanalys, varför valet av rätt verktyg och noggrann dokumentation av koden var avgörande för framgången.

Verktyg som Används i Projektet:

Datainsamling och Lagring:

SQL Databaser: För lagring och hantering av insamlade data användes SQL-baserade databaser, vilket möjliggjorde effektiv hantering och frågeställning av stora datamängder.

IoT Plattformer: För att samla in data från olika källor i den industriella miljön användes olika IoT-plattformar.

Dataförädling och Analys:

Python: Som huvudspråk för dataanalys användes Python, tack vare dess mångsidighet och rika ekosystem av datavetenskapsbibliotek.

Pandas & NumPy: För datamanipulering och beräkningar användes Pandas och NumPy, vilket möjliggjorde effektiv hantering av stora datamängder.

Jupyter Notebooks: För interaktiv utveckling och dokumentation användes Jupyter Notebooks, vilket underlättade processen att dela och granska kod.

Versionskontroll och Samarbete:

Git & GitHub: För versionskontroll och samarbete kring kodutveckling användes Git, med GitHub som plattform för kodlagring och delning.

Kod Dokumentation:

Kod Kommentarer: Koden dokumenterades noggrant med kommentarer för att förklara funktioner, variabler och logik, vilket underlättade förståelsen och underhållet av koden.

Funktion Dokumentation: Varje funktion och metod skrevs med tydliga beskrivningar, parameterförklaringar och exempel på användning..

Dokumentation av Kod för Prediktivt Underhåll Baserat på Drivdata

Översikt:

Denna Python-kod använder maskininlärning (specifikt en Random Forest-klassificerare) för att förutse underhållsbehov baserat på data om energiförbrukning och vridmoment (torque) från industriell utrustning. Koden inkluderar dataförädling, funktionstillverkning, modellträning och utvärdering.

Kodbibliotek:

pandas för datahantering.

numpy för numeriska operationer.

sklearn för maskininlärningsfunktioner, inklusive datadelning, förbehandling och modellutvärdering.

Kodflöde och Funktioner:

Importera Data:

Använder pandas för att läsa in en CSV-fil som innehåller driftdata.

Datarengöring:

Korrigerar negativa värden i 'Energy Consumption (kWh)' genom att ta absolutvärden, vilket säkerställer att energiförbrukningsdata är positiva.

Introducera Anomalier:

Skapar syntetiska anomalier i 1% av datasetet för 'Torque (Nm)' och 'Energy Consumption (kWh)' för att simulera oväntade driftsförhållanden.

Funktionstillverkning:

Beräknar rullande medelvärden för vridmoment och energiförbrukning för att fånga trender över tid.

Normalisering av Data:

Använder MinMaxScaler från sklearn.preprocessing för att skala utvalda funktioner till ett intervall mellan 0 och 1, vilket förbättrar prestandan hos många maskininlärningsmodeller.

Förbereda Tränings- och Testdata:

Delar upp data i funktioner (X) och målvariabel (y).

Använder train_test_split för att dela data i tränings- och testuppsättningar.

Modellträning:

Tränar en Random Forest-klassificerare på träningsdatan.

Modellutvärdering:

Använder testdata för att göra förutsägelser och utvärderar modellens prestanda med konfusionsmatris, klassificeringsrapport och noggrannhetsvärde.

Funktionens Viktighet:

Presenterar en lista över funktioners viktighet enligt modellen, vilket hjälper till att förstå vilka funktioner som mest bidrar till modellens förutsägelser.

Spara Modellen (Valfritt):

Inkluderar en kommenterad kod för att spara den tränade modellen för framtida användning med joblib.

Användningsinstruktioner:

Innan du kör koden, se till att 'Drive_data_Jan-Feb.csv' finns tillgänglig och innehåller korrekta data.

Koden kan modifieras för att passa olika dataset genom att ändra filnamnet eller kolumnnamnen.

Om modellen ska sparas, avkommentera sista delen och se till att joblib är installerat.

Underhåll och Utvidgning:

För att utöka funktionaliteten kan ytterligare funktioner skapas och utvärderas, eller andra maskininlärningsmodeller kan testas för jämförelse av prestanda.