大学共同利用機関法人 情報・システム研究機構
国立情報学研究所
National Institute of Informatics

トップエスイー ソフトウェア開発実践演習

AIを活かすためのデータ前処理方法の検討

日立製作所　伊藤　渉太
NECソリューションイノベーション　大河内　駿太郎
日本総合研究所　清野宗一郎

開発における問題点

今後AI人材の不足状況はますます深刻化していく見通しであり、AI人材確保には社内育成が第一選択である。しかしながら、現況は環境が整っておらず、中でもAIの前処理は初学者にとってはハードルが高いものである。

手法・ツールの適用による解決

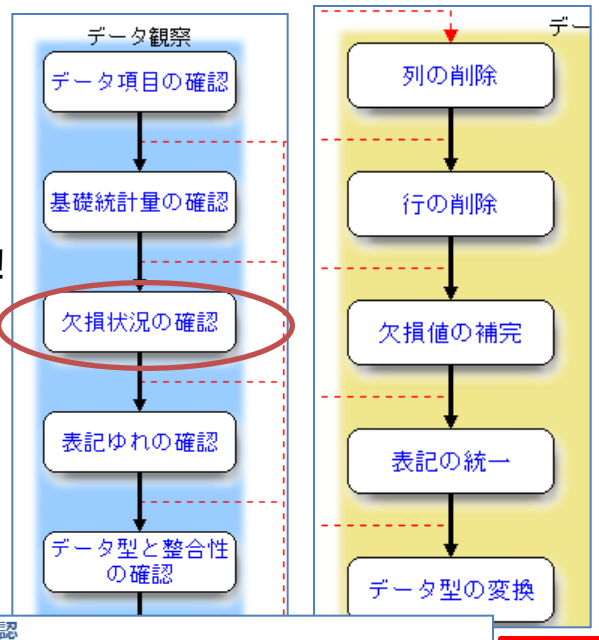
ここでは，データ分析初学者をターゲットとしてペルソナ設定およびカスタマージャーニーマップによる分析を行うことで、フレームワークを作成し、初学者の生産性を向上し、作業の品質を一定に引き上げることを目指した。

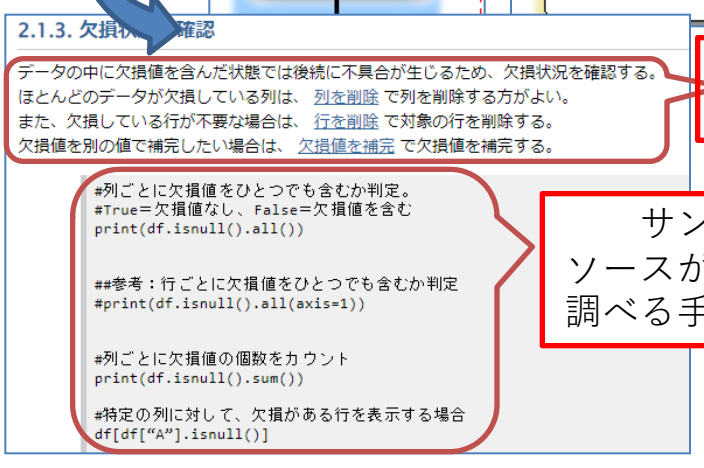
フレームワークの概要

フレームワークは以下の内容で構成
- データ分析の流れ
- 前処理（データ観察・データ加工）のフローチャート
- データ観察作業の詳細（サンプルコード、判断基準）
- データ加工作業の詳細（サンプルコード、判断基準）

初学者がどんな作業を実施すべきか分かりやすい

クリック！！





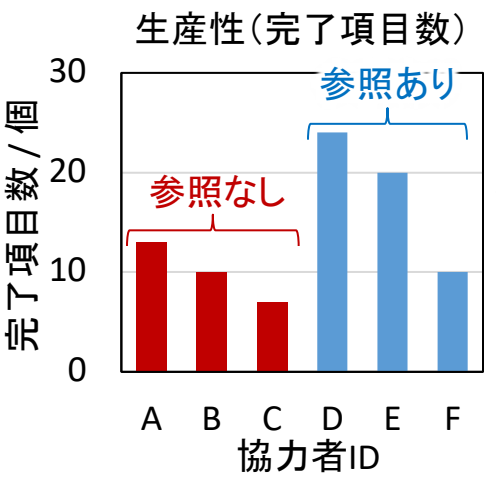
次の作業が分かる！

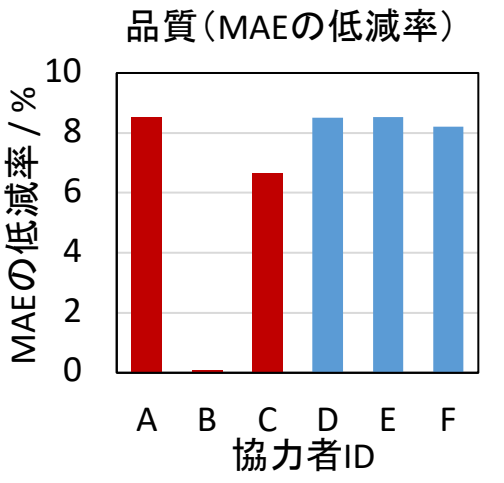
サンプルソースがあるから調べる手間が不要

フレームワークの効果検証(条件)

評価条件	概要
データ	・気象庁の地点気象データセット
予測対象	・翌日の東京の平均気温
評価対象	・データ解析の初学者6名（Python使用経験はあり）
評価方法	・評価対象をフレームワーク参照可否の2グループに分割し、2時間の前処理を依頼
評価指標	・完了項目数（生産性） ・平均絶対誤差（MAE）の低減率（品質）

フレームワークの効果検証(結果)





独立2群のt検定（分散は等しいと仮定）

項目	結果	平均の有意差
生産性	0.0066 (P値) < 0.05 (有意水準)	あり
品質	0.2657 (P値) > 0.05 (有意水準)	ありとは断定できない

フレームワークの参照により生産性が有意に向上