20240812 Ext. ミーティング議事録

**日時**: 2024年8月12日 10:00～11:30

**場所**: オンライン会議 (Zoom)

**参加者**:

- クライアント側: 田中一郎 (プロジェクトマネージャー), 鈴木花子 (品質保証担当), 佐藤次郎 (データエンジニア)

- データ分析チーム: 山田太郎 (プロジェクトリーダー), 中村美咲 (データサイエンティスト), 小林健 (データエンジニア), ChatGPT (アシスタント)

アクションアイテム:

* データ分析チーム: データの前処理を一週間以内に完了 (担当: 中村美咲, 小林健)
* クライアント側: 追加のデータが必要な場合は速やかに提供 (担当: 佐藤次郎)
* 山田太郎: 初期分析結果を8月26日のミーティングで共有

決定事項:

* 次回ミーティング: 2024年8月26日 10:00～11:00

議題 1: 修理データの現状把握

- 田中一郎: クライアント側から提供された修理データの概要を説明。データは過去5年間に渡る修理履歴、部品交換履歴、および故障の原因分析に関する情報を含む。

- データ量: 約10GB

- データフォーマット: CSVファイル形式

- 含まれる情報: 製品ID、修理日、修理内容、使用部品、修理担当者、故障原因など。

- 鈴木花子: 修理データは品質保証の観点からも重要であり、製品の品質改善に役立つと期待している。特に、頻繁に発生する故障やリコールに繋がる問題を特定したいとのこと。

議題 2: データ前処理の方針

- 山田太郎: 提供されたデータの前処理について議論。まずはデータのクリーニング作業を実施し、不正確なデータや欠損値の補完を行う必要がある。

- データのクリーニング手順として、重複データの削除、欠損値の補完方法の検討、異常値の確認などを提案。

- 中村美咲: データのクリーニングに続いて、特徴量エンジニアリングを行い、分析に必要なデータを抽出・変換する予定。

- 特に、故障原因や修理内容に基づいた新たな特徴量を作成し、モデルの精度向上を図る。

議題 3: 分析の進め方

- 小林健: 分析の第一ステップとして、データの可視化を行い、データの全体像を把握することが重要。これにより、データに潜むパターンや異常値を発見しやすくなる。

- 次のステップとして、製品カテゴリごとの故障原因の頻度分析を提案。

- 山田太郎: これらの結果をもとに、予測モデルを構築する予定。特に、どの製品がどの程度のリスクを持つかを予測するモデルが重要。

- 使用する手法として、ランダムフォレストや勾配ブースティングを考慮中。

議題 4: 今後のスケジュール

- 田中一郎: クライアント側として、初期分析結果を2週間以内に共有してほしいとの要望があり、その後に更なるデータ分析の方向性を決定したいとのこと。

- 次回の進捗確認ミーティングは8月26日に予定。

- 山田太郎: スケジュールを承認。データの前処理を一週間以内に完了し、初期分析結果を取りまとめて報告する計画。

以上