経済産業省 商務情報政策局 産業保安グループ 製品安全課 御中

令和2年度 経済産業省デジタルプラットフォーム構築事業
製品安全分野におけるデータ利活用調査

最終報告

アクセンチュア株式会社
2021年3月26日（金）

accenture

目次

1. 調査の概要
- 1-1. 調査の背景・目的
- 1-2. スケジュール
- 1-3. サマリ

2. 傾向分析の成果
- 2-1. 作業の全体像
- 2-2. クローリング技術検証
- 2-3. データ収集
- 2-4. データクレンジング
- 2-5. 法令違反等の確認
- 2-6. データ分布状況の確認
- 2-7. 仮説選定・検定
- 2-8. 機械学習モデル構築・精度検証

3. 予兆分析の成果
- 3-1. 作業の全体像
- 3-2. 検索データ収集・分析結果
- 3-3. SNSデータ収集・分析結果
- 3-4. 予兆検知手法検証
- 3-5. ダッシュボード作成

4. 運用に向けた考察
- 4-1. 傾向分析の活用とモデル運用
- 4-2. 予兆分析ダッシュボードの運用
- 4-2. ネットパトロール効率化

5. 運用に向けた課題・対応策
6. 第三者有識者コメント

Copyright © 2020 Accenture All rights reserved

1. 調査の概要

Copyright © 2020 Accenture All rights reserved

目次

1. 調査の概要
- 1-1. 調査の背景・目的
- 1-2. スケジュール
- 1-3. サマリ

2. 傾向分析の成果
- 2-1. 作業の全体像
- 2-2. クローリング技術検証
- 2-3. データ収集
- 2-4. データクレンジング
- 2-5. 法令違反等の確認
- 2-6. データ分布状況の確認
- 2-7. 仮説選定・検定
- 2-8. 機械学習モデル構築・精度検証

3. 予兆分析の成果
- 3-1. 作業の全体像
- 3-2. 検索データ収集・分析結果
- 3-3. SNSデータ収集・分析結果
- 3-4. 予兆検知手法検証
- 3-5. ダッシュボード作成

4. 運用に向けた考察
- 4-1. 傾向分析の活用とモデル運用
- 4-2. 予兆分析ダッシュボードの運用
- 4-2. ネットパトロール効率化

5. 運用に向けた課題・対応策
6. 第三者有識者コメント

Copyright © 2020 Accenture All rights reserved

事業の背景

✓ 「インターネット取引が増加するなか、インターネット上で販売される製品における違反件数も増加傾向にあり、製品安全の確保に向け様々な取組みが行われている

製品安全の確保に向けた取組み※1

製品事故の未然防止策

✓ **製品安全関係法※2による事前規制**

- 技術基準体系、規制対象品目の見直し
- 規制対象製品における技術基準適合義務、PSマーク表示義務　など

製品事故の再発防止策

✓ **重大製品事故情報の収集・公表**

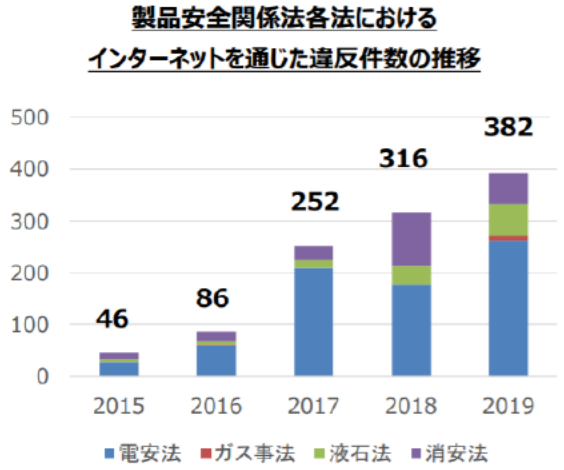
✓ **消費者への情報提供**

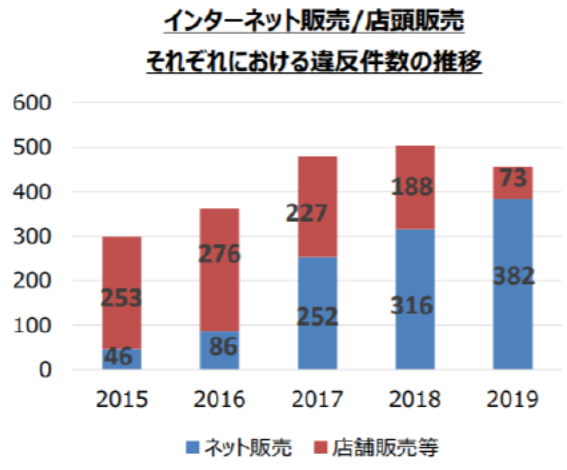
インターネット取引への対応強化

✓ **危険な製品販売の自粛を働きかけ**

- 製品安全関係法に違反する製品のほか、重大製品事故発生のリスクが高いことが判明した製品の出品停止　など

インターネット取引等における製品安全関係法違反件数の現状※3





※1: 経済産業省「インターネットを通じた違反品販売の現状」、「製品安全政策の概要と動向について」、「インターネット取引における製品安全に関する提言」
※2：消費生活用製品安全法、電気用品安全法、ガス事業法、液化石油ガスの保安の確保及び取引の適正化に関する法律の４法を指す
※3: 経済産業省「2019年の製品安全に関する動向」 資料4-3　インターネット取引における製品安全中のスライド３を引用
https://www.meti.go.jp/product_safety/policy/2019fyreport.html

Copyright © 2020 Accenture All rights reserved

本業務の目的

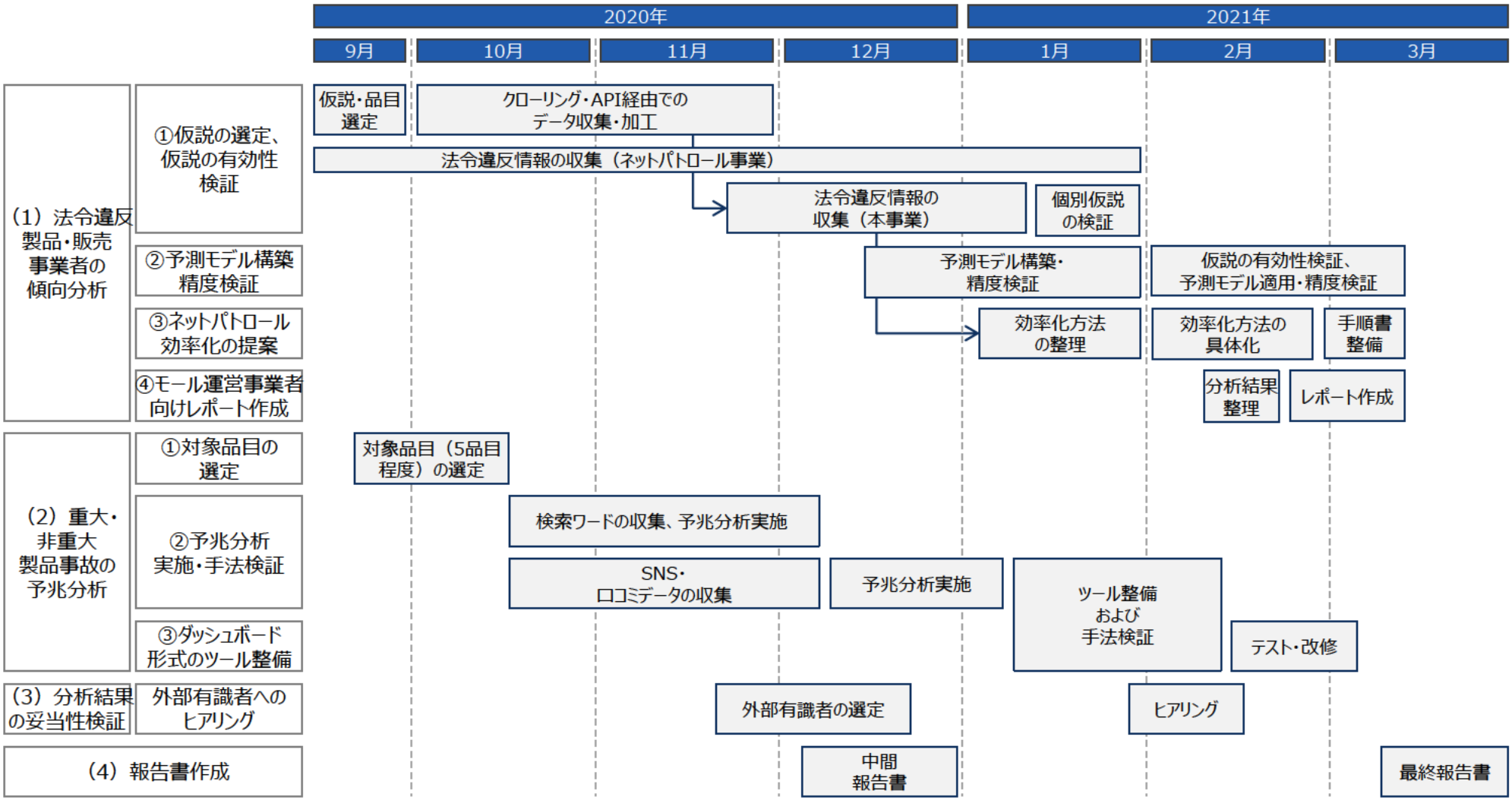
本調査では、インターネット取引における法令違反・製品事故リスクをデータから抽出し、製品事故の防止や被害の拡大防止に活用すべく、分析方法及び活用方法を提案し貴省業務の高度化を目指す

		実施内容
（1）法令違反製品・販売事業者の傾向分析	①仮説（15件）の有効性検証	• ３つのモール運営事業者に出品される重点5品目についてクローリング等により情報収集し、法令違反傾向に係る15仮説を統計検定 • データのクレンジングの効率化・運用に向けた示唆だし
	②分析結果の予測精度検証	• ①で収集したデータを用い、法令違反傾向を予測するモデルを構築し精度検証を実施（モール運営事業者横断、品目横断含む）
	③ネットパトロール効率化の提案	• ①および②の成果をネットパトロール事業への活用方法を提案
	④モール運営事業者への連携を想定したレポート作成	• モール運営事業者ごとに、法令違反事業者の傾向を可視化
（2）重大・非重大製品事故の予兆分析	①対象品目（5品目程度）の選定	• 検索データ、SNS・口コミデータの件数も確認した上で選定
	②検索ワード、SNS・口コミデータから予兆分析実施・手法検証	• 履歴推移から時系列での予兆を把握
	③ダッシュボード形式のツールイメージ作成	• 画面イメージを作成（Excelなど）
（3）分析結果の妥当性検証	外部有識者（3名程度）の提案、ヒアリング実施	• 統計学の学識経験者、データ分析実務者により分析および機械学習モデル構築のアプローチ等に関するアドバイス・評価

Copyright © 2020 Accenture All rights reserved

1. 調査の概要　1-2. スケジュール

作業スケジュール



Copyright © 2020 Accenture All rights reserved

本レポートのサマリとの対応（傾向分析）

３つのECサイトに出品される重点**5品目**の収集データおよび**昨年度分析データ**および**ネットパトロール事業データ**を用いて**15仮説に対する仮説検証**と**モデル構築・精度検証**を実施

□仮説検証：

※　法令違反データが少ないため分析結果の解釈に留意が必要

- ✓ 一部の仮説において、商品および販売事業者の特徴と法令違反傾向との間に関連性が認められた。その傾向は、ECサイト、品目ごとに異なる場合があることが明らかとなった
- ✓ 法令違反の予備軍と考えられる「照会後商品削除」傾向についてもあわせて分析し、複数の法令違反リスク傾向が明らかとなった
- ✓ 有効な仮説を組み合わせた活用方法を提案

□モデル構築・精度検証：

※　法令違反データが少ないためモデル精度等の解釈に留意が必要

- ✓ 法令違反例のサンプル数が少ないため、品目横断での予測モデルのみ構築し、精度評価を実施（照会後商品削除傾向を予測するモデルについても別途構築・精度検証）

法令違反予測精度概要：　※　照会後商品削除予測も別途あわせて実施

- 品目横断モデル：適合率（最高）100%、再現率（最高）約55%

- ✓ 因子ごとに影響度の高低が認められた
- ✓ モデル構築・精度検証から得られた示唆の活用方法を提案

□今後に向けた課題：

- ✓ 継続した法令違反データ収集と検証の更新
- ✓ アンサンブル等のモデルの発展
- ✓ 法令違反確認（PSマーク表示確認）をより効果的・効率的に実施する仕組みづくり

Copyright © 2020 Accenture All rights reserved

本レポートのサマリとの対応（予兆分析）

インターネット検索エンジンにおける検索データおよびSNSデータ等を収集し、製品事故の予兆を捉えることができるかを分析・検証を実施した。その結果を踏まえ、予兆を捉えるためのツールイメージを作成

□予兆分析：

- ✓ 検索データを分析し、「品番」×「事故ワード」等の検索量により予兆検知の可能性を複数発見
- ✓ SNSデータと検索データを含め予兆検知の可能性を分析し、「品番又はメーカー名が含まれる投稿」をシグナルとして用いる予兆検知方法を検証
- ✓ 上記を実現するツールイメージを作成

□今後に向けた課題：

- ✓ 品番一覧・メーカー名一覧の拡充
- ✓ SNSデータのクリーニング精度向上
- ✓ 継続したデータ収集と運用ノウハウの拡充

Copyright © 2020 Accenture All rights reserved

2.傾向分析の成果

Copyright © 2020 Accenture All rights reserved

目次

1. 調査の概要
- 1-1. 調査の背景・目的
- 1-2. スケジュール
- 1-3. サマリ

2. 傾向分析の成果
- 2-1. 作業の全体像
- 2-2. クローリング技術検証
- 2-3. データ収集
- 2-4. データクレンジング
- 2-5. 法令違反等の確認
- 2-6. データ分布状況の確認
- 2-7. 仮説選定・検定
- 2-8. 機械学習モデル構築・精度検証

3. 予兆分析の成果
- 3-1. 作業の全体像
- 3-2. 検索データ収集・分析結果
- 3-3. SNSデータ収集・分析結果
- 3-4. 予兆検知手法検証
- 3-5. ダッシュボード作成

4. 運用に向けた考察
- 4-1. 傾向分析の活用とモデル運用
- 4-2. 予兆分析ダッシュボードの運用
- 4-2. ネットパトロール効率化

5. 運用に向けた課題・対応策
6. 第三者有識者コメント

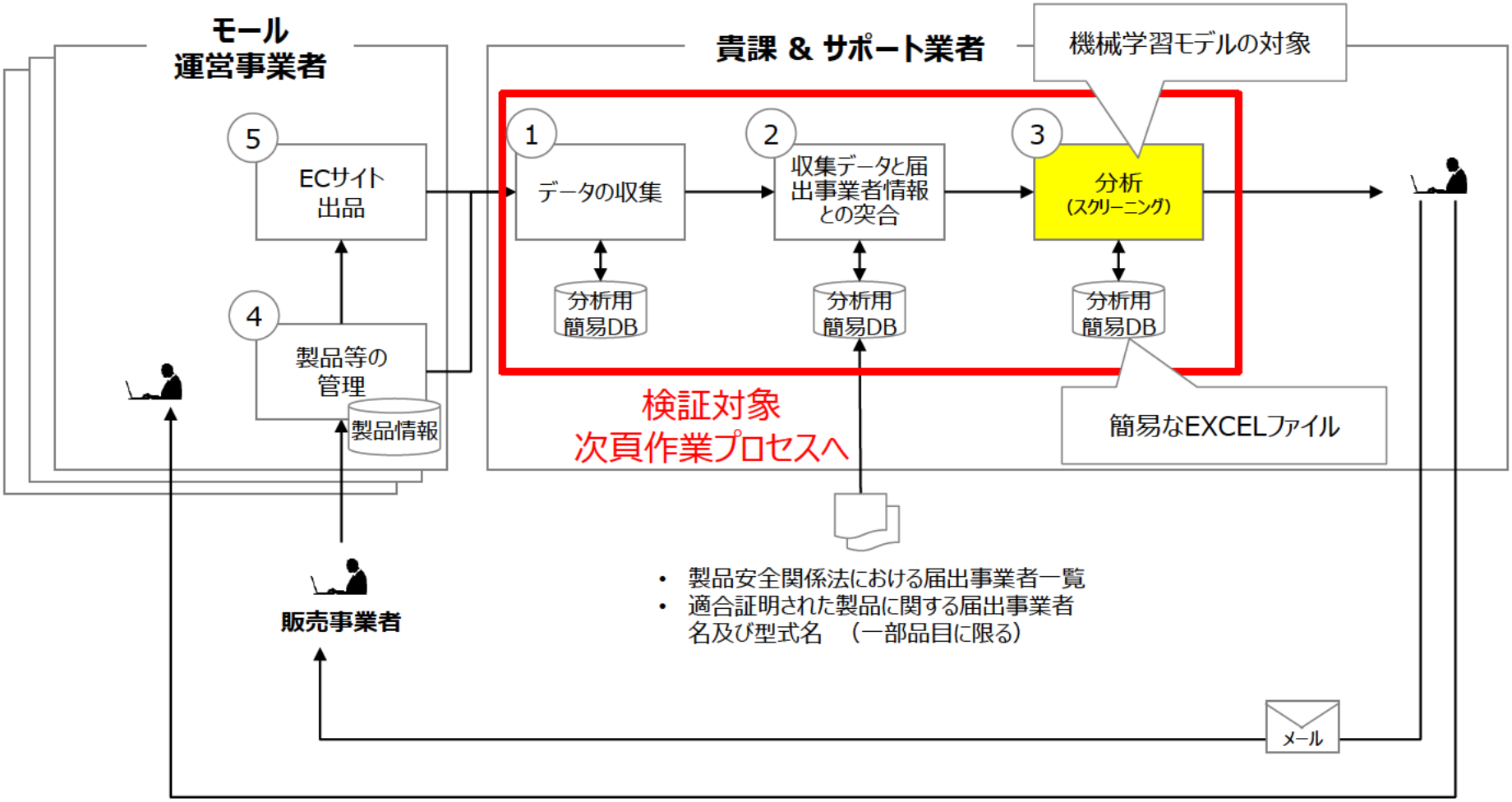
Copyright © 2020 Accenture All rights reserved

2. 傾向分析の成果　2-1. 作業の全体像

今後の業務活用イメージ

傾向分析の結果と機械学習モデルにより貴課業務の効率化を目指す

（将来像）分析結果の活用イメージ



Copyright © 2020 Accenture All rights reserved

2. 傾向分析の成果　2-1. 作業の全体像

プロセスと本書の構成の対応関係

✓ ①～⑥の流れで調査・検証を実施

1 データ収集およびデータクレンジング

- ✓ 分析に必要な情報がデータ収集可能か技術検証する
- ✓ webクローリング等により製品・販売者情報を収集する
- ✓ 販売者の情報については、モール運営事業者と情報連携
- ✓ 収集したデータのノイズを除去し分析に適した形に整える

2 法令違反ラベル付与

- ✓ 販売事業者に対しPSマーク表示の有無をメールにて問合せ、回答状況に応じて経産省確認の上、法令違反データとする

3 データ分布状況の確認

- ✓ 母集団と標本集団の分布の差異を評価

4 仮説検証

- ✓ 15仮説に対してPSマーク表示および照会後商品削除行動との関連性を確認
- ✓ 統計検定を実施

5 法令違反予測モデル構築・精度検証

- ✓ 勾配ブースティング等複数の機械学習を活用しモデル構築および精度検証を実施

6 有識者ヒアリング

- ✓ 第三者有識者へのヒアリングを実施し分析結果の妥当性および改善点についてコメントをいただく

※ 分析においてはPSマーク表示（法令違反関係）に限らず、商品ページ削除に至った案件も分析した。

Copyright © 2020 Accenture All rights reserved

2. 傾向分析の成果

プロセスと本書の構成の対応関係

1 データ収集およびデータクレンジング

- ✓ 分析に必要な情報がデータ収集可能か技術検証する
- ✓ webクローリング等により製品・販売者情報を収集する
- ✓ 販売者の情報については、モール運営事業者と情報連携
- ✓ 収集したデータのノイズを除去し分析に適した形に整える

2 法令違反ラベル付与

- ✓ 販売事業者に対しPSマーク表示の有無をメールにて問合せ、回答状況に応じて経産省確認の上、法令違反データとする

3 データ分布状況の確認

- ✓ 母集団と標本集団の分布の差異を評価

4 仮説検証

- ✓ 15仮説に対してPSマーク表示および照会後商品削除行動との関連性を確認
- ✓ 統計検定を実施

5 法令違反予測モデル構築・精度検証

- ✓ 勾配ブースティング等複数の機械学習を活用しモデル構築および精度検証を実施

6 有識者ヒアリング

- ✓ 第三者有識者へのヒアリングを実施し分析結果の妥当性および改善点についてコメントをいただく

Copyright © 2020 Accenture All rights reserved

WebクローリングおよびAPIによる収集の技術的有用性

✓ 分析に必要な商品データと販売事業者データをWebクローリングおよびAPIによりデータ収集可能であることを確認

詳細は次「2-4.データクレンジング」に記載

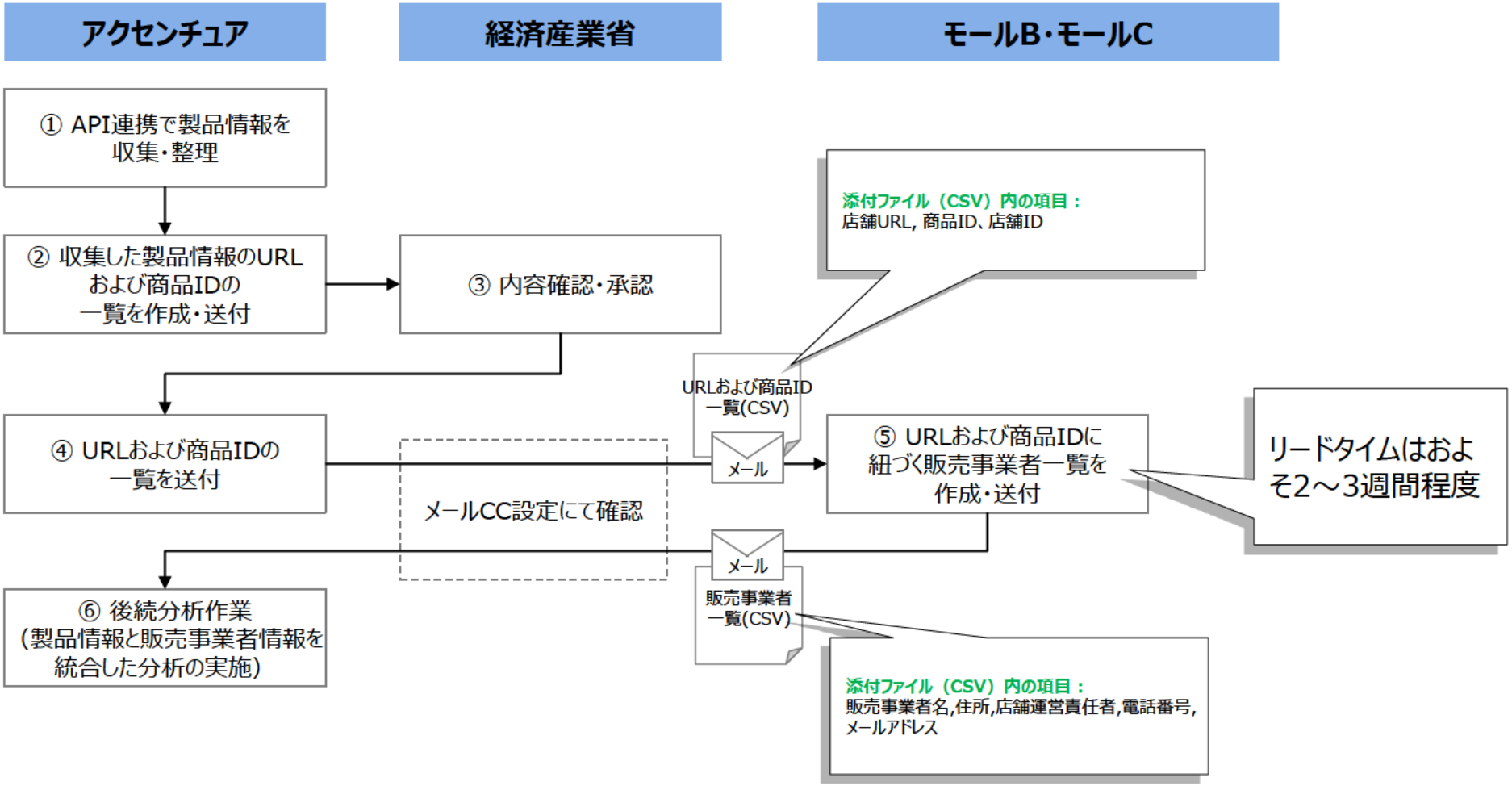
	本取組みでのデータ取得方法		今後の運用に向けた検証	
	商品データ	販売事業者データ	検証内容	検証結果
モールA	クローリング	クローリング	クローリングで取得した両データが大きく欠けずに紐付けられるか	商品データの86.4%に販売事業者データが紐付けられ、運用上十分な件数を確保
モールB モールC	API	モール運営事業者へ依頼	販売事業者データのクローリングによる取得が可能か	本取組みにて必要な販売事業者データの取得が技術的可能と確認

- 本検証ではPython言語によりWebクローリングおよびAPIをコールしデータ収集する検証用プログラムを作成した。
- 実際に運用する場合は相応の技術者によりツールを開発するとともに、商品ページ構成の変更にあわせツールをメンテナンス・維持（改修含む）するか、技術者がいない場合はモール運営事業者に協力を要請しデータ取得する方策が考えられる。

Copyright © 2020 Accenture All rights reserved

販売事業者情報の紐づけプロセス（モールB・モールC）

- 収集した商品データに対して以下のプロセスにより販売事業者情報との紐づけを全件実施
- モールAについてはWebクローリングにより取得・紐づけを実施



Copyright © 2020 Accenture All rights reserved

データクレンジング状況

- ✓ 以下のプロセスによりデータクレンジング処理（ノイズ除去）作業を実施
- ✓ 前提として各モールの商品カテゴリを用いて絞り込み済みの状態から開始

データクレンジング作業

テキストベース

1．商品定義からの抽出ワードによる対象外商品の除外

2．商品一覧からの抽出ワードによる対象外商品の除外※

画像ベース（サンプリング）

3．商品画像による対象外商品の除外
（画像による絞込み）

※：具体的には「内蔵ケース」等のキーワードにより、「バッテリーそのものではなく梱包ケースを販売している場合」や異なる商品を販売しているものを除外する。

Copyright © 2020 Accenture All rights reserved

データクレンジング後のデータ件数

- ✓ 製品安全4法対象品目の内、指定５品目の情報を収集およびクレンジング実施
- ✓ データ収集およびクレンジングにより77,178件の商品および販売事業者情報を取得

収集データ数（ノイズ除去※）

対象品目（5品目）	モールA (クローリング取得)	モールB (API取得)	モールC (API取得)	合計
バッテリー・モバイルバッテリー	4,734	17,640	5,207	27,581
カートリッジガスこんろ	1,617	4,112	3,989	9,718
乗車用ヘルメット	5,620	6,025	1,212	12,857
電気温風機	6,117	4,650	1,222	11,989
LEDランプ	5,628	5,357	4,048	15,033
				77,178

※ノイズ除去とはWebクローリング等で収集したデータの中に、対象品目とは異なる製品（ノイズ）が混在していた場合にデータから除外する作業を意図

Copyright © 2020 Accenture All rights reserved

データクレンジング作業を踏まえた運用向け示唆

- ✓ 今後の運用ではテキストベースによるデータクレンジング作業を定型化することが可能と考える
- ✓ 5品目に対して2名体制で2週間程度で定型化可能ではないか
- ✓ なお完全なデータクレンジング（ノイズ除外）はできないため法令遵守有無確認の際は商品ページ確認が必要

	データクレンジング作業	今後の運用に向けた示唆
テキストベース	1．商品定義からの抽出ワードによる対象外商品の除外	定型化が可能であり、ECサイトのカテゴリを用いることでノイズを除去できる
テキストベース	2．商品一覧からの抽出ワードによる混在する対象外商品の除外	経産省のノウハウ（商品知識や出品者の行動）を活かしプラスαの定型化が可能ではないか
画像ベース（サンプリング）	3．商品画像による対象外商品の除外（画像による絞込み）	稼働がかかり、テキストベースである程度の精度は保たれるため、必須ではない（照会の際は事前確認が必要）

※ データクレンジング以外ではあるが、画像確認を通じて得られた示唆として販売事業者ごとにPSマーク画像等のページ内配置位置に特徴/傾向があり販売事業者別に画像確認作業を進めると作業効率が上がるものと考える

Copyright © 2020 Accenture All rights reserved

2. 傾向分析の成果

プロセスと本書の構成の対応関係

1 データ収集およびデータクレンジング

- ✓ 分析に必要な情報がデータ収集可能か技術検証する
- ✓ webクローリング等により製品・販売者情報を収集する
- ✓ 販売者の情報については、モール運営事業者と情報連携
- ✓ 収集したデータのノイズを除去し分析に適した形に整える

2 法令違反ラベル付与

- ✓ 販売事業者に対しPSマーク表示の有無をメールにて問合せ、回答状況に応じて経産省確認の上、法令違反データとする

3 データ分布状況の確認

- ✓ 母集団と標本集団の分布の差異を評価

4 仮説検証

- ✓ 15仮説に対してPSマーク表示および照会後商品削除行動との関連性を確認
- ✓ 統計検定を実施

5 法令違反予測モデル構築・精度検証

- ✓ 勾配ブースティング等複数の機械学習を活用しモデル構築および精度検証を実施

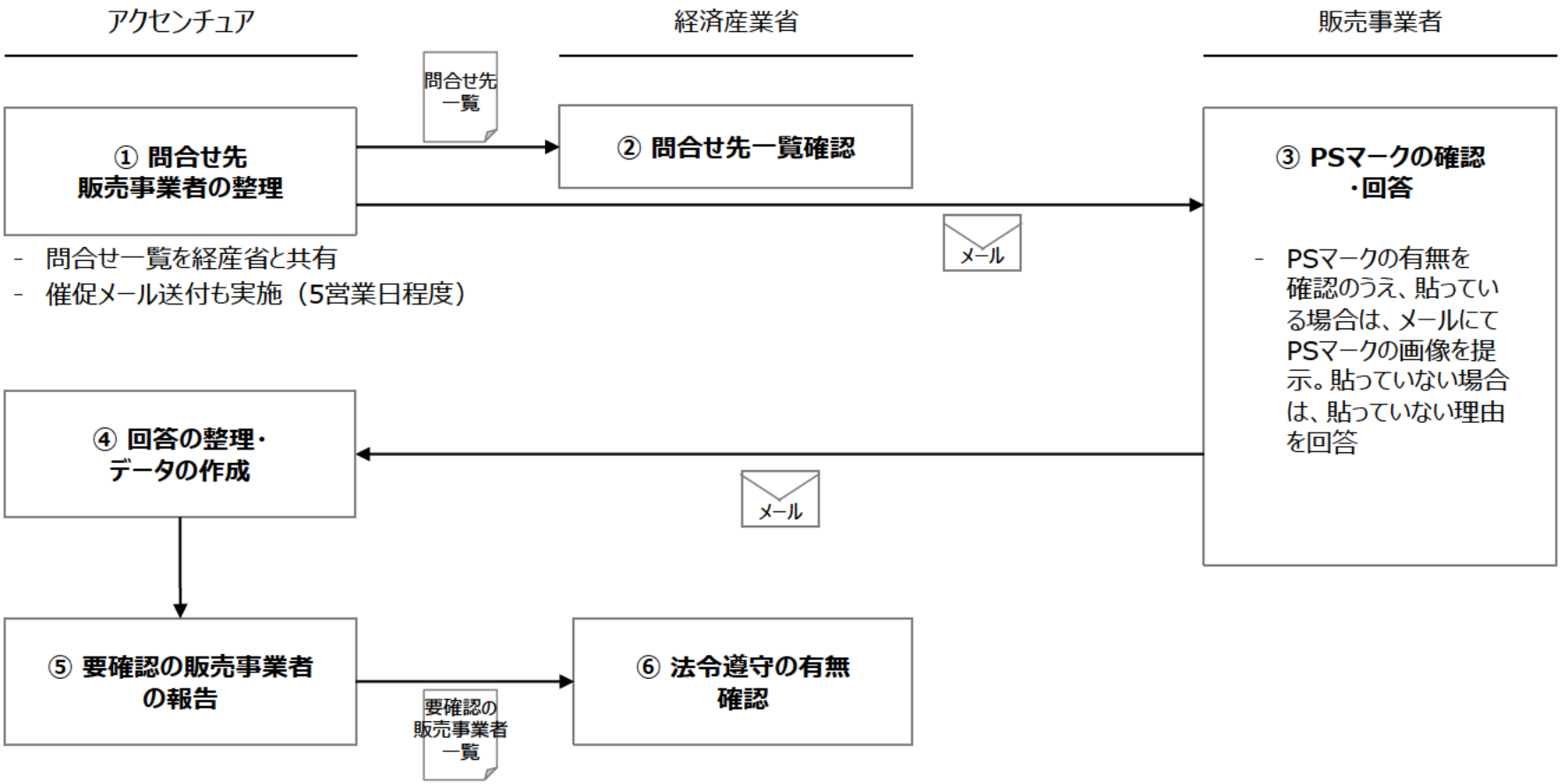
6 有識者ヒアリング

- ✓ 第三者有識者へのヒアリングを実施し分析結果の妥当性および改善点についてコメントをいただく

Copyright © 2020 Accenture All rights reserved

販売事業者への法令遵守確認プロセス

✓ 販売事業者からの回答のうち、PSマーク画像の提示がないものについて経済産業省による法令違反の有無を確認



Copyright © 2020 Accenture All rights reserved

2. 傾向分析の成果

プロセスと本書の構成の対応関係

1 データ収集およびデータクレンジング

- ✓ 分析に必要な情報がデータ収集可能か技術検証する
- ✓ webクローリング等により製品・販売者情報を収集する
- ✓ 販売者の情報については、モール運営事業者と情報連携
- ✓ 収集したデータのノイズを除去し分析に適した形に整える

2 法令違反ラベル付与

- ✓ 販売事業者に対しPSマーク表示の有無をメールにて問合せ、回答状況に応じて経産省確認の上、法令違反データとする

3 データ分布状況の確認

- ✓ 母集団と標本集団の分布の差異を評価

4 仮説検証

- ✓ 15仮説に対してPSマーク表示および照会後商品削除行動との関連性を確認
- ✓ 統計検定を実施

5 法令違反予測モデル構築・精度検証

- ✓ 勾配ブースティング等複数の機械学習を活用しモデル構築および精度検証を実施

6 有識者ヒアリング

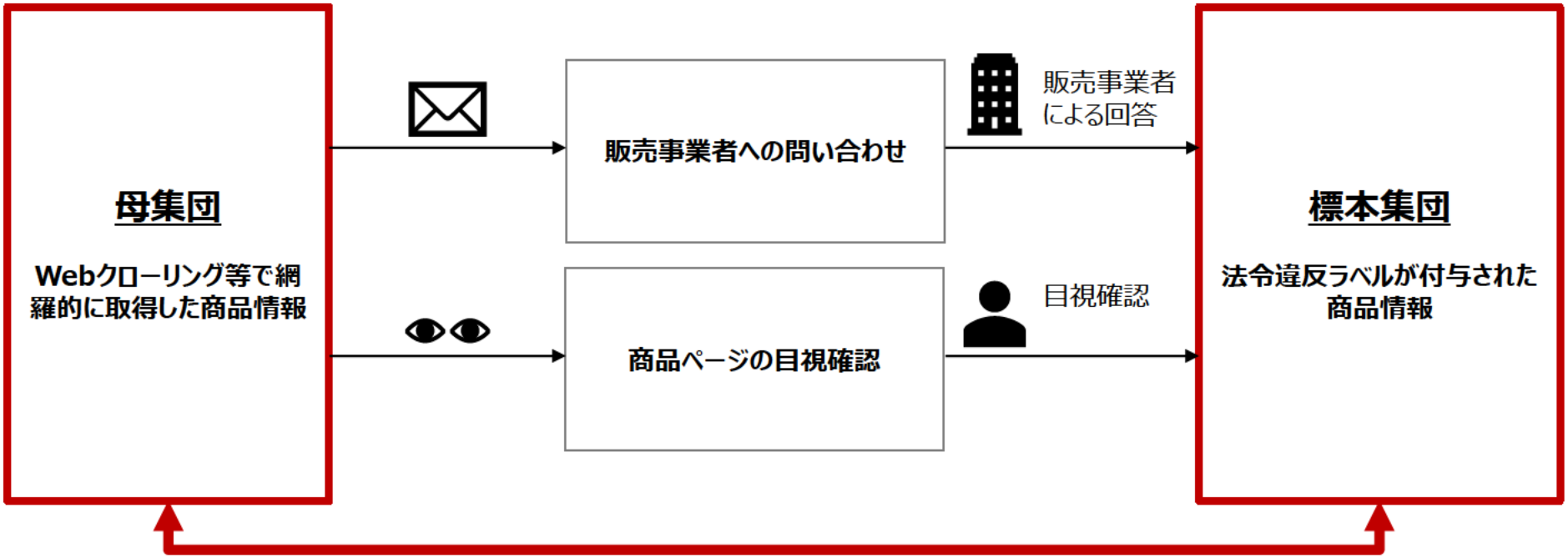
- ✓ 第三者有識者へのヒアリングを実施し分析結果の妥当性および改善点についてコメントをいただく

Copyright © 2020 Accenture All rights reserved

分布の確認

✓ Webクローリング等により取得した商品群（母集団）と法令違反ラベルを付与した商品群（標本集団）の分布傾向を確認し、集団間の差異を評価

母集団と標本集団との対応と「法令違反の判定の流れ」および比較対象の集団



母集団と標本集団の分布傾向の差異を確認

✓ 分布傾向に大きな差異が認められる項目については分析時の層別因子として考慮

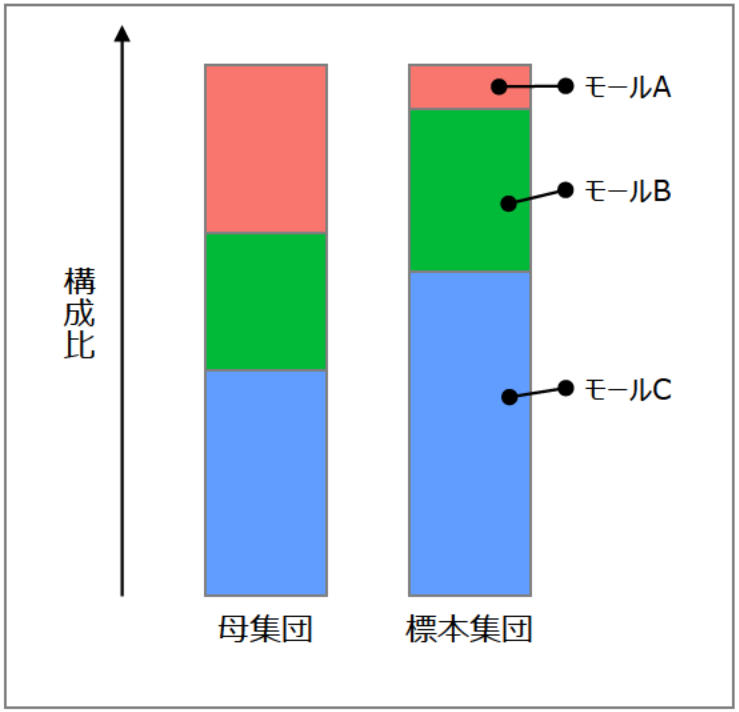
Copyright © 2020 Accenture All rights reserved

分布傾向の比較の流れ

✓ 母集団（Webクローリング等により取得した商品群）と標本集団（法令違反ラベルを付与した商品群）の分布傾向の違いを仮説検定および構成比の差分の観点から評価

分布傾向の比較の流れ

母集団と標本集団の分布



（比較軸をモール運営事業者とした例）

仮説検定による分布の比較

母集団と標本集団における比較軸の構成比に差異があるかを評価
（カイ二乗検定）

有意水準10%

構成比の差分による比較

母集団と標本集団における比較軸の構成比の差の大きさを評価
（例えば、母集団のモールAの割合と標本集団のモールAの割合の差分を評価）

Copyright © 2020 Accenture All rights reserved

2. 傾向分析の成果

プロセスと本書の構成の対応関係

1 データ収集およびデータクレンジング

- ✓ 分析に必要な情報がデータ収集可能か技術検証する
- ✓ webクローリング等により製品・販売者情報を収集する
- ✓ 販売者の情報については、モール運営事業者と情報連携
- ✓ 収集したデータのノイズを除去し分析に適した形に整える

2 法令違反ラベル付与

- ✓ 販売事業者に対しPSマーク表示の有無をメールにて問合せ、回答状況に応じて経産省確認の上、法令違反データとする

3 データ分布状況の確認

- ✓ 母集団と標本集団の分布の差異を評価

4 仮説検証

- ✓ 15仮説に対してPSマーク表示および照会後商品削除行動との関連性を確認
- ✓ 統計検定を実施

5 法令違反予測モデル構築・精度検証

- ✓ 勾配ブースティング等複数の機械学習を活用しモデル構築および精度検証を実施

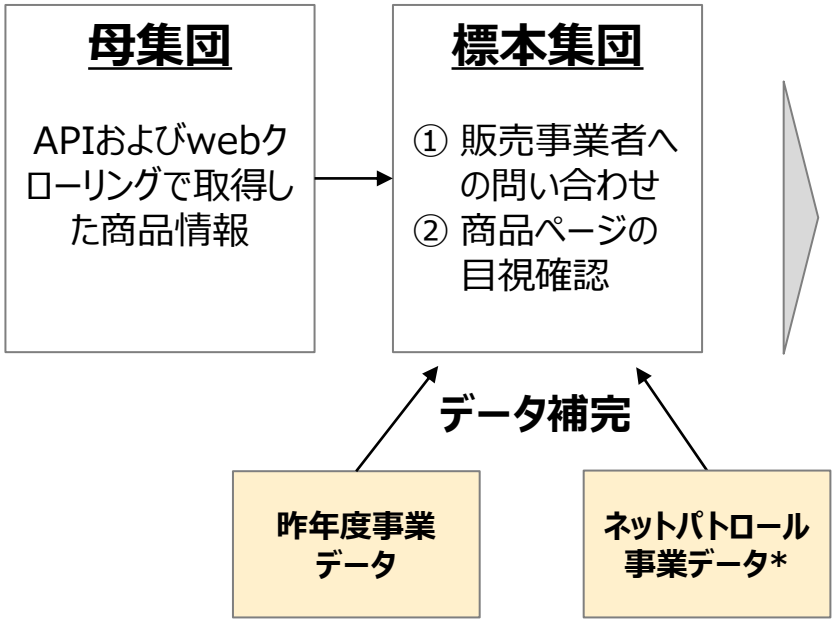
6 有識者ヒアリング

- ✓ 第三者有識者へのヒアリングを実施し分析結果の妥当性および改善点についてコメントをいただく

Copyright © 2020 Accenture All rights reserved

データ補完と照会後商品削除例の取り扱い

- ✓ 今年度事業データは法令違反例が少ないため、**昨年度事業にて取得済のデータおよびネットパトロール事業データを追加**して分析を実施
- ✓ 「照会後商品削除」を分析対象外と扱う分析（分析パターン1）、「照会後商品削除」を法令違反に分類する分析（分析パターン2）をそれぞれ実施



法令違反ラベル	照会後商品削除例の扱い	
	分析パターン1	分析パターン2
法令適合	法令適合	法令適合
照会後商品削除	分析対象外	法令違反
法令違反	法令違反	

＊：ネットパトロール事業とは本事業と並行した経済産業省によるECサイト上の法令違反が疑われる商品への調査事業

Copyright © 2020 Accenture All rights reserved

仮説一覧

✓ 経済産業省における仮説から重要度および収集データの充足度等を踏まえて15個を検証仮説として選出

検証仮説

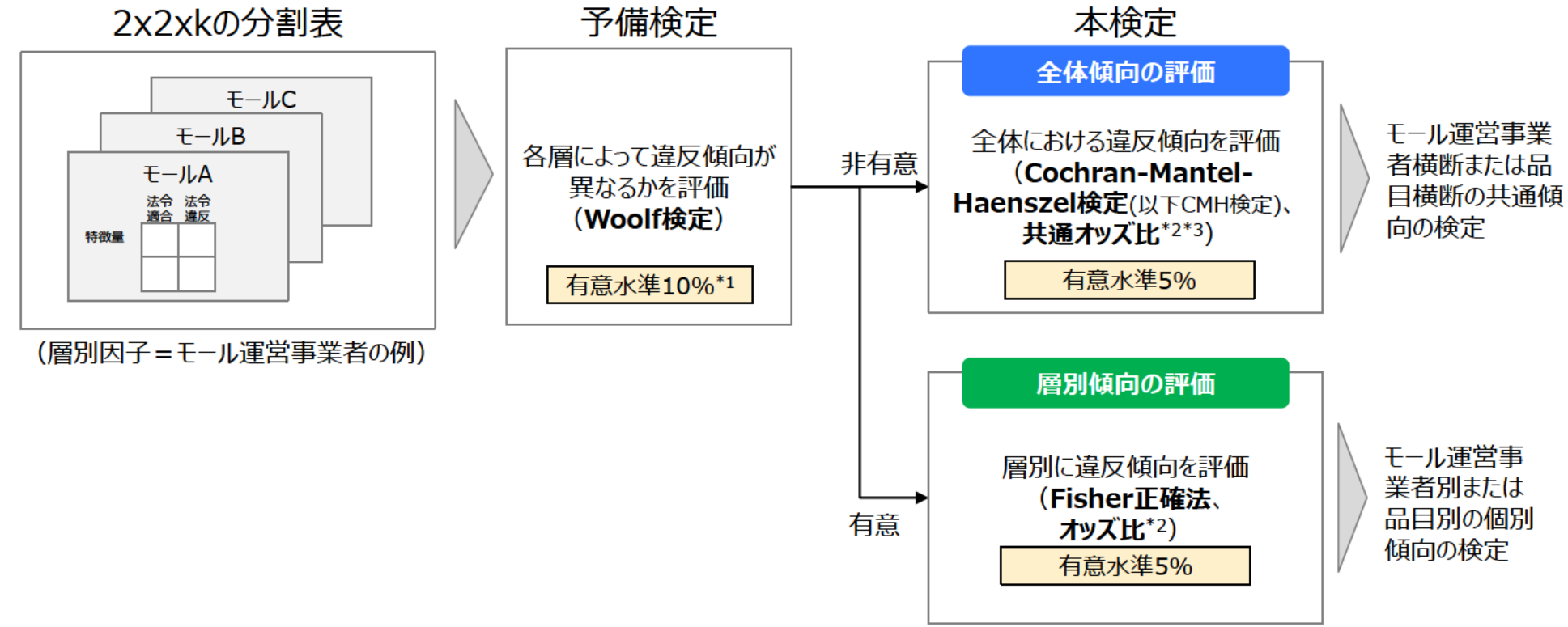
商品	仮説1	商品概要における情報量が少ないと法令違反リスクが高い
	仮説2	購入者レビューが少ないと法令違反リスクが高い
	仮説3	購入者評価点数が低いと法令違反リスクが高い
	仮説4	代金支払い方式が限定的であると法令違反リスクが高い
	仮説5	商品の属性（品目タイプ等）ごとに法令違反リスクが異なる
	仮説6	商品説明に「並行輸入」「インポート」が含まれると法令違反リスクが高い
	仮説7	商品説明に「PSマーク」の単語があると、法令違反リスクが低い
	仮説8	在庫が無いと法令違反リスクが高い
	仮説9	配送期間が長い（即日配送の記載がない）と法令違反リスクが高い
	仮説10	価格が異常に高い・低いと法令違反リスクが高い
販売事業者	仮説11	販売事業者が外国住所であると、法令違反リスクが高い
	仮説12	販売元が他に販売している商品が低評価であると法令違反リスクが高い
	仮説13	販売事業者評価件数が少ないと法令違反リスクが高い
	仮説14	販売事業者評価点数が低いと法令違反リスクが高い
	仮説15	販売事業者の店舗数により法令違反リスクが異なる（販売事業者の規模観点）

Copyright © 2020 Accenture All rights reserved

仮説検証の流れ

- ✓ 母集団と標本集団での分布傾向の差異が大きい「モール運営事業者」「品目」を層別因子として考慮する
- ✓ 層別因子（モール運営事業者、品目）それぞれについて、各層の違反傾向に違いがあるかを評価し（予備検定）、違いが認められない場合は全体での傾向として評価し、層間での差異が認められる場合は、層別に仮説検定を実施する

仮説検証の流れ



*1:統計有識者へのヒアリング結果を踏まえて、予備検定であるWoolf検定の有意水準を保守的に設定

*2:分割表において頻度0のセルがある場合は、Haldane-Anscombe補正した修正オッズ比を計算

*3:共通オッズ比は、層別した複数の2x2表を統合した各層に共通なオッズ比

Copyright © 2020 Accenture All rights reserved

仮説検証結果での用語説明

項番	記号／用語	説　明
1	Woolf検定	予備検定で用いた手法。当検定で「有意差あり」となった仮説は「違反傾向が層ごとに異なる」ため「モール運営事業者個別」および「品目個別」に検定（Fisher正確法）する。一方、当検定で「有意差なし」となった仮説は「違反傾向が層ごとに異ならない」ため横断的に検定（CMH検定）する。
2	CMH検定	Woolf検定結果が「有意差なし」の場合に、「モール運営事業者横断」および「品目横断」での検定方法。当検定で「有意差あり」となった仮説は、「モール運営事業者横断」および「品目横断」での業務効率化・高度化に活用し得る事項。
3	Fisher正確法	Woolf検定結果が「有意差あり」の場合に、「モール運営事業者個別」および「品目個別」での検定方法。当検定で「有意差あり」となった仮説は、「モール運営事業者個別」および「品目個別」に特化した形で業務効率化・高度化に活用し得る事項。

Copyright © 2020 Accenture All rights reserved

オッズ比を用いる理由

- ✓ 臨床研究におけるケースコントロール研究（後ろ向き研究）の考え方を採用
- ✓ 仮説検証の評価指標の１つとしてオッズ比を用いる

$$\text{オッズ比} = \frac{p1/(1-p1)}{p2/(1-p2)} = \frac{A \times D}{B \times C}$$

※ p1 = A/(A+B),　1-p1 = B/(A+B), p2 = C/(C+D), 1-p2 = D/(C+D)

説明変数例	法令違反	法令適合	合計
商品レビュー3点以上	A	B	A+B
商品レビュー3点未満	C	D	C+D
合計	A+C	B+D	A+B+C+D

※ AからDは各セルのサンプル数を表す
※ 法令違反率pが小さい場合はオッズ比はリスク比の近似となる
※ 選択バイアス（サンプリングに偏り）が小さい場合はリスク比が用いられる

Copyright © 2020 Accenture All rights reserved

2. 傾向分析の成果

プロセスと本書の構成の対応関係

1 データ収集およびデータクレンジング

- ✓ 分析に必要な情報がデータ収集可能か技術検証する
- ✓ webクローリング等により製品・販売者情報を収集する
- ✓ 販売者の情報については、モール運営事業者と情報連携
- ✓ 収集したデータのノイズを除去し分析に適した形に整える

2 法令違反ラベル付与

- ✓ 販売事業者に対しPSマーク表示の有無をメールにて問合せ、回答状況に応じて経産省確認の上、法令違反データとする

3 データ分布状況の確認

- ✓ 母集団と標本集団の分布の差異を評価

4 仮説検証

- ✓ 15仮説に対してPSマーク表示および照会後商品削除行動との関連性を確認
- ✓ 統計検定を実施

5 法令違反予測モデル構築・精度検証

- ✓ 勾配ブースティング等を複数の機械学習を活用しモデル構築および精度検証を実施

6 有識者ヒアリング

- ✓ 第三者有識者へのヒアリングを実施し分析結果の妥当性および改善点についてコメントをいただく

Copyright © 2020 Accenture All rights reserved

法令違反予測モデルの構築　分析アプローチ

✓ 勾配ブースティング等のアルゴリズムを用いて、法令違反を予測するモデル構築し、予測結果に影響を与える要因について評価する

分析対象データ

予測したい指標（目的変数）*1

法令違反

X

特徴量（説明変数）

- 商品の属性
- 商品評価
- 支払い
- 配送方法
- 販売事業者の属性
- 取り扱い商品情報
- 事業者レビュー

等

分析フロー

データ分割 → 学習 → 精度評価

学習データ
(全体の70%データを用いてCross-Validationを実施)

検討アルゴリズム
・勾配ブースティング (lightGBM, XGBoost)
・Random Forest
・ロジスティック回帰

評価データ
(全体の30%)

モデル精度の評価
特徴量の影響度
について評価

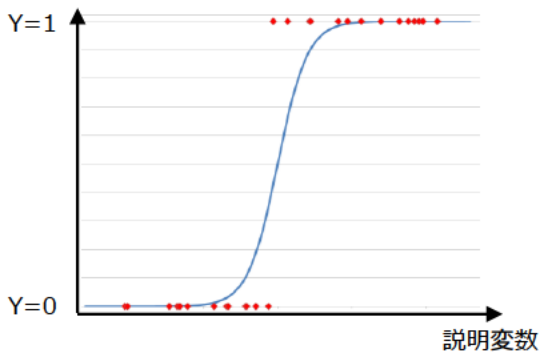
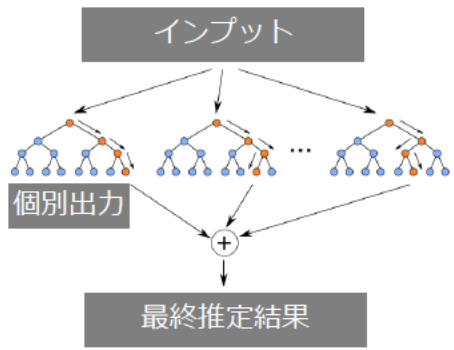
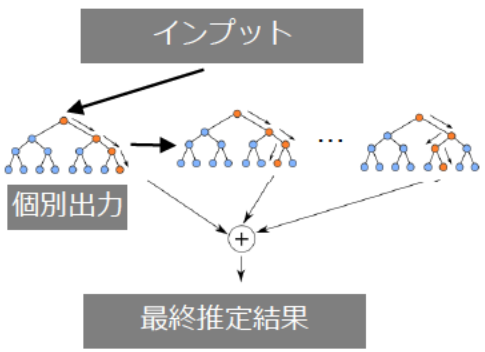
*1：「照会後商品削除」例を法令違反として扱う分析（分析パターン1）
および「照会後商品削除」例を法令適合として扱う分析（分析パターン2）

Copyright © 2020 Accenture All rights reserved

検討アルゴリズム

✓ 複数の機械学習アルゴリズムを用いて、違反傾向を予測するモデルを構築

別々に学習させて作成した**複数のモデルを融合**させることによって、
よりよい予測精度を目指すメタアルゴリズムの手法

	回帰分析	アンサンブル学習	
		バギング	ブースティング
	ロジスティック回帰	Random Forest	勾配ブースティング (lightGBM, XGBoost)
イメージ			
概要	• **回帰分析の一種**で目的変数（違反の有無）と説明変数（特徴量）の関係性を評価する	• **独立した複数のモデル**を作成する • 特定のモデルで誤りが存在した場合でも、複数の独立したモデルで**多数決で判断**することで、予測精度を向上させる	• **非独立な複数のモデル**を作成する • 複数のモデルを作成するにあたって、**前回のモデルの誤りを修正するようなモデルを追加**していくことで、予測精度を向上させる

Copyright © 2020 Accenture All rights reserved

評価指標

✓ 混同行列を用い、モデルによる予測値と実際の値を比較することで、モデルを評価する

混同行列

件数	実際の結果 Yes/**違反あり**	実際の結果 No/**違反なし**
予測 Yes（違反**あり**と予測）	①真陽性 (TP:True Positive)	②偽陽性 (FP:False Positive)
予測 No（違反**なし**と予測）	③偽陰性 (FN:False Negative)	④真陰性 (TN:True Negative)

混同行列を用いた指標

正解率 (Accuracy)	$\frac{TP + TN}{TP + FP + FN + TN}$	• 全ての判定対象のうち、正しく予測された件数
誤分類率 (Misclassification rate)	$\frac{FP + FN}{TP + FP + FN + TN}$	• 全ての判定対象のうち、誤って予測された件数
適合率 (Precision)	$\frac{TP}{TP + FP}$	• Yesと予測されたもののうち、正しく予測された件数
再現率 (Recall)	$\frac{TP}{TP + FN}$	• 実際にYESだったもののうち、正しくYesと予測できた件数
F値	$2 \times \frac{適合率 \times 再現率}{適合率 + 再現率}$	• 適合率と再現率の調和平均で両者のバランスをとった指標

使用分野や予測目的によって重要視される指標は異なる（以下例）

- クレジットカードの不正使用のように、一つの不正が大きな損失につながる場合・・再現率を重要
- 開封確率や購入確率の高い層に向けて効果的にDM送付をしたい場合・・・適合率を重視

※各指標とも0～1の範囲の値を取る
※※誤分類率を除き、値が高いほど良い指標

Copyright © 2020 Accenture All rights reserved

検討ステップ

- ✓ まずは、品目に依存しない品目横断での予測モデルを構築し、法令違反に共通的に認められる要因について考察
- ✓ 次に、品目別に予測モデルを構築し、品目横断モデルとの差異を考察し、品目特有の要因について把握する

検討ステップ

品目横断モデルの作成	品目別モデルの作成
対象とする品目が増えるごとに法令違反予測モデルを一から作成するのは実施上困難であるため、まずは、品目横断での予測モデルを構築し、**品目に依存しない法令違反のリスク要因を把握**する	品目横断モデルでは捉えきれない、**品目特有の要因について把握**する

Copyright © 2020 Accenture All rights reserved

2. 傾向分析の成果　2-8. 機械学習モデル構築・精度検証

検討ステップ

分析パターン1
法令違反

分析パターン2
法令違反(照会後商品削除含む)

分析パターン

①法令違反	②法令違反（照会後商品削除含む）
法令違反となる販売事業者の特徴について評価する	照会後に商品ページの削除・販売中止するなど法令遵守意識のついて疑義の生じる販売事業者の特徴について評価する

✖

検証ステップ

品目横断モデルの作成	品目別モデルの作成
対象とする品目が増えるごとに法令違反予測モデルを一から作成するのは実施上困難であるため、まずは、品目横断での予測モデルを構築し、**<u>品目に依存しない法令違反のリスク要因を把握</u>**する	品目横断モデルでは捉えきれない、**<u>品目特有の要因について把握</u>**する

Copyright © 2020 Accenture All rights reserved

2. 傾向分析の成果　2-8. 機械学習モデル構築・精度検証

予測モデル　サマリ①

✓ **品目横断モデルの**構築・精度評価を現状のデータにより実施

法令違反予測精度概要：

- 品目横断モデル：最高適合率100%、最高再現率54.5%

※ 法令違反サンプルが少ないため、モデル精度評価等の解釈に注意する必要がある

分析結果

（共通的な傾向 品目横断）	• 因子ごとに影響度の高低が認められた
品目別モデルの傾向	• サンプル数が少ないため品目別モデルは作成せず

今後の発展

✓ **<u>データ取得の効率化</u>**
モール運営事業者と連携し、PSマーク掲示のルール化など、効率的な法令違反サンプルの収集の仕組みを検討

✓ **<u>品目横断モデルの成長</u>**
法令違反サンプルを更に収集し、品目横断モデルを再学習させモデルの成長を図る

✓ **<u>品目別モデルの構築</u>**
更なるデータ収集を進めながら、個々の品目の特性に特化したモデルを構築し更なる精度向上を図る

Copyright © 2020 Accenture All rights reserved

検討ステップ

分析パターン	①法令違反	②法令違反（照会後商品削除含む）
	法令違反となる販売事業者の特徴について評価する	照会後に商品ページの削除・販売中止するなど法令遵守意識のついて疑義の生じる販売事業者の特徴について評価する

✖

検証ステップ	品目横断モデルの作成	品目別モデルの作成
	対象とする品目が増えるごとに法令違反予測モデルを一から作成するのは実施上困難であるため、まずは、品目横断での予測モデルを構築し、**品目に依存しない法令違反のリスク要因を把握**する	品目横断モデルでは捉えきれない、**品目特有の要因について把握**する

Copyright © 2020 Accenture All rights reserved

2. 傾向分析の成果　2-8. 機械学習モデル構築・精度検証

予測モデル　サマリ②

✓ **品目横断モデル**および**品目別モデル（重点５品目：バッテリー・モバイルバッテリー、カートリッジガスこんろ、乗車用ヘルメット、電気温風機、LEDランプ）**の構築・精度評価を現状のデータにより実施

予測精度概要：
- 品目横断モデル：適合率約95%、再現率約70%
- 品目別モデル：適合率70～100%、再現率40～100%（品目により異なる）

分析結果

（共通的な傾向 品目横断）
- 因子ごとに影響度の高低が認められた

品目別モデルの傾向
- 品目別に固有の因子が確認された
- 品目ごとに傾向が異なり固有の性質/傾向があることが伺える

今後の発展

✓ **<u>データ取得の効率化</u>**
モール運営事業者と連携し、PSマーク掲示のルール化など、法令違反データの収集を容易にする

✓ **<u>モデルの組み合わせ活用</u>**
データ収集を更に進めながら、品目横断モデルと品目別モデルとを組み合わせたアンサンブル学習*等を用いて精度向上を図ることも考えられる

* 複数のモデル（学習器）を融合させ学習モデルを生成する手法であり、モデル構造が複雑化するが精度向上を見込める可能性がある

Copyright © 2020 Accenture All rights reserved

2. 傾向分析の成果

まとめ

	サマリ	実施項目
（1）法令違反製品・販売事業者の傾向分析	Webクローリングによるデータ収集の技術的有用性を示すとともにデータクレンジングについての運用示唆をまとめ、15個の仮説検証および機械学習モデル構築・精度検証を実施し、実運用に向けた示唆を提示した。	✓ データ収集・データクレンジング完了 ✓ 法令違反確認完了 ✓ 15仮説の検証完了 ✓ 機械学習モデル構築・精度検証完了 ✓ ネトパト事業連携実施完了
（2）重大・非重大製品事故の予兆分析	検索データによる予兆検知の可能性を複数発見、SNSデータを含めて予兆検知の可能性を分析し、「品番またはメーカー名が含まれる投稿」をシグナルとして用いる予兆検知方法を提示した。 上記を実現するツールイメージも作成した。	✓ 検索データ収集・検証完了 ✓ SNSデータ収集完了 ✓ SNSデータ分析完了 ✓ ダッシュボードツールイメージ作成完了
（3）分析結果の妥当性検証	第三者有識者３名にヒアリングを実施し分析アプローチ等の妥当性を確認した。	✓ 有識者３名調整・案件説明完了 ✓ レポート提示・ヒアリング完了

Copyright © 2020 Accenture All rights reserved

3.予兆分析の成果

Copyright © 2020 Accenture All rights reserved

目次

1. 調査の概要
- 1-1. 調査の背景・目的
- 1-2. スケジュール
- 1-3. サマリ

2. 傾向分析の成果
- 2-1. 作業の全体像
- 2-2. クローリング技術検証
- 2-3. データ収集
- 2-4. データクレンジング
- 2-5. 法令違反等の確認
- 2-6. データ分布状況の確認
- 2-7. 仮説選定・検定
- 2-8. 機械学習モデル構築・精度検証

3. 予兆分析の成果
- 3-1. 作業の全体像
- 3-2. 検索データ収集・分析結果
- 3-3. SNSデータ収集・分析結果
- 3-4. 予兆検知手法検証
- 3-5. ダッシュボード作成

4. 運用に向けた考察
- 4-1. 傾向分析の活用とモデル運用
- 4-2. 予兆分析ダッシュボードの運用
- 4-2. ネットパトロール効率化

5. 運用に向けた課題・対応策
6. 第三者有識者コメント

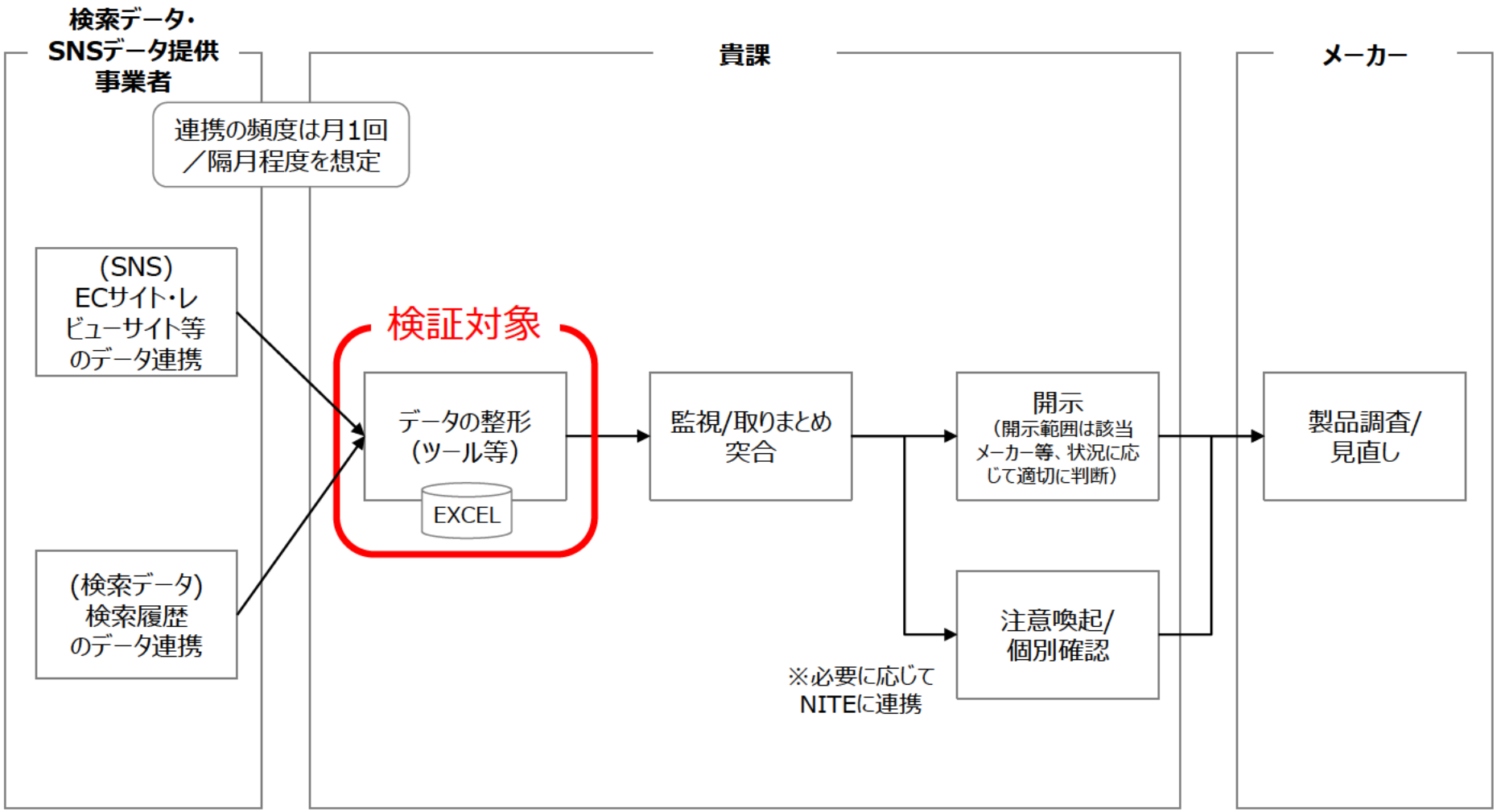
Copyright © 2020 Accenture All rights reserved

3. 予兆分析の成果　3-1. 作業の全体像

今後の各データの活用プロセスイメージ

SNSデータおよび検索データを隔週または月１回程度の頻度で取得し、監視・注意喚起等を行う

（将来像）分析結果の活用イメージ



Copyright © 2020 Accenture All rights reserved

検証項目

業務運用に向けて適切なワードや想定工数の確認、分析結果の信頼性について検証する。

	検証方針	検証項目	検証内容
昨年度検証からの拡大	1 検索データ+SNSデータによる検出品目の幅拡大 SNSデータによる製品特定の拡大余地を検証	「**品目** × 製品事故ワード」での出現	✓ 様々な性質のデータでの出現結果を比較し、「品目」単位の出現確認に適したデータを確認
昨年度検証からの信頼性強化	2 検索データ×SNSデータによる証跡補強（信頼性向上） 検索データ･SNSデータの組み合わせによる分析結果の信頼性を検証	「品目 × **メーカー名や製品型番** × 製品事故ワード」での出現	✓ 製品特定可能な品目を確認 ✓ 品目や製品事故ワードに応じた各種データの使い分けを検証 ✓ 適切な製品事故ワード等を確認
		「**リコール、重大事故製品型番** × 製品事故ワード」での事故発生前の出現	✓ 重大事故発生日の前後における「事故製品型番」の出現を確認 ✓ 事故報告のない「製品型番」での出現を確認･他データでの結果と比較 ✓ 業務運用に向けた想定工数を確認
昨年度検証からの拡大	3 画像データ等による製品特定の可能性を検証	**画像･動画**を含むデータによる製品特定	✓ 画像･動画を含む製品事故データの出現を確認 ✓ 画像･動画による製品特定を品目別に確認

Copyright © 2020 Accenture All rights reserved

参考：検証利用データ収集要件

✓ 検索データおよびSNSデータのデータ収集要件を示す

データ種別	目的	条件	収集期間
検索データ	**仮説①：**具体的な事故事象を示す**「品目」**と**「製品事故ワード」**の組み合わせ検索の有無・変化を確認する **仮説②：**（仮説①の検索を確認した後）「製品型番」など製品を特定可能な情報を確認する	• 「品目」×「事故ワード」の検索スコア※ • 「品目」×「製品型番」×「事故ワード」の検索スコア※	2016/12/1-2019/11/30
SNSデータ	**仮説：**検索データと組み合わせて予兆を捉えることができるかメーカーおよび製品型番を含む投稿等により事故予兆を検知可能か確認する	• 「事故ワード」の投稿データ	2016/11/30-2020/11/30

※ 検索スコアとは「『品目(バッテリー)×製品事故ワード』を含むキーワードで実際に検索された回数に、定数を積算してスコア化（マスキング）したもの」

Copyright © 2020 Accenture All rights reserved

検証プロセス

✓ 検索データとSNSデータについてそれぞれ以下のプロセスで実施

データ収集・加工

- **対象品目と事故関連ワード**でデータ収集し、クレンジング実施（特に、SNSデータは不要データが多く含まれる可能性があるため、注意する）
- 重大・非重大事故の**報告事例をピックアップ**し、事故報告の**数年前からの投稿推移を確認**するべくデータ抽出

データ可視化、予兆分析実施

- 過去の重大・非重大事故事例を題材に、**投稿数の推移を可視化**
- 予兆分析に必要な**指標の定義**、及び**しきい値の設定**について検討
- **メーカー名**もしくは**品番の頭文字**との関連性をみることで具体的な**製品の特定**が可能か、という点についても検討

予兆分析の手法検討（抽出条件定義）

- 収集したデータの**クレンジング**から、投稿数の推移グラフの**考察ポイント**、具体的な**製品特定の方法**等について、手順を整理し、**簡易的なマニュアル**を作成

Copyright © 2020 Accenture All rights reserved

検索データ収集・分析サマリ

- ✓ 検索データ収集・分析を実施
- ✓ 品目によって状況は異なるが製品型番による予兆検知（仮説②）の可能性を発見

予兆可能性あり
予兆確認できず

品目		仮説① 「品目」×「製品事故ワード」の組み合わせ検索の有無・変化があるか	仮説② （①の後）「メーカー名」「製品型番」など製品を特定可能な情報がないか	後続検証作業
対象品目	バッテリー	事故近辺に変化事例ありだが予兆とは言えず（傾向はモバイルバッテリーに一致）	事故前**検索あり** 品番×事故ワード検索142件/3年	SNSデータを含めた仮説検証
	モバイルバッテリー	事故近辺に変化事例ありだが予兆とは言えず	事故前**検索あり** 品番×事故ワード検索32件/3年	
	携帯電話機	事故前検索増加は確認できず	事故前**検索あり** 品番×事故ワード検索48件/3年	
	ノートパソコン	事故前検索増加は確認できず	事故前検索確認できず 品番×事故ワード検索292件/3年	
	電機ストーブ	事故前検索増加は確認できず （事例はあるが季節変動の可能性あり）	事故前**検索あり** 品番×事故ワード検索8件/3年	
	照明器具	事故前検索増加は確認できず	品番含む検索なし （商品特性によるものと考えられる）	

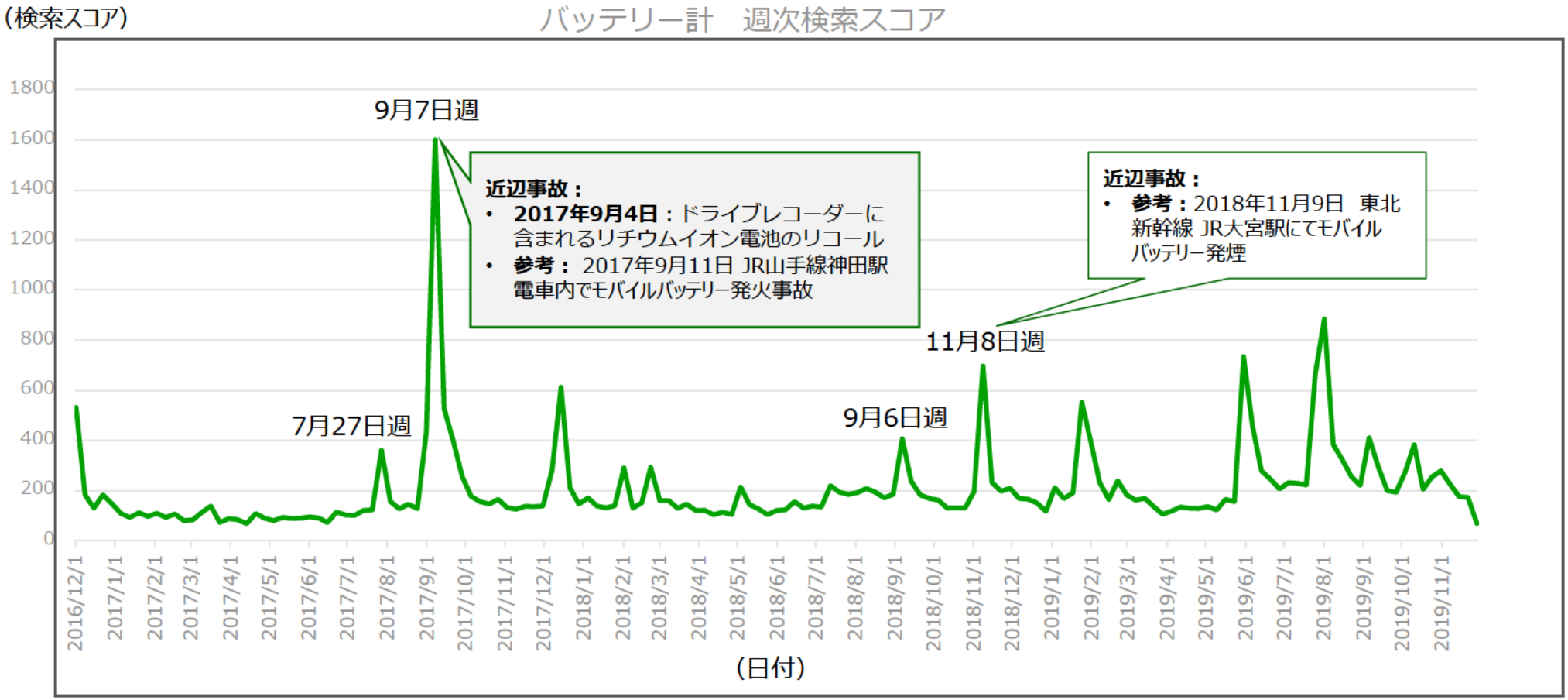
Copyright © 2020 Accenture All rights reserved

3. 予兆分析の成果　3-2. 検索データ収集・分析結果

バッテリー

検索データの分析経過　～バッテリーおよびモバイルバッテリー　品目出現

- ✓ 仮説②「品番×製品事故ワード」の組み合わせにて事故予兆の可能性を見出し、詳細を分析（事故前の検索の増加傾向を確認したものの表中の事故に関して製品型番などによる直接的な関連性は確認できなかった）
- ✓ バッテリー検索の傾向はモバイルバッテリーとほぼ一致（バッテリー検索の8割をモバイルバッテリーが占めるため）



※ 検索スコアとは「『品目(バッテリー)×製品事故ワード』を含むキーワードで実際に検索された回数に、定数を積算してスコア化（マスキング）したもの」

Copyright © 2020 Accenture All rights reserved

収集するSNSデータ

SNSデータ収集においては各ECサイト、レビューサイト等の情報を利用する

	検証方針	検証目的	データの要件
1	SNSデータによる製品特定の拡大余地を検証	製品特定可能な品目を拡大することで分析可能な領域を拡大したい	✓ 特になし
2	検索データ・SNSデータの組み合わせによる分析結果の信頼性を検証	事故リスクが高いとメーカーに示せるなどより実効力のある分析結果とする	✓ 単語ベースの検索データよりも情報の詳細度が高いデータ
3	画像データ等による製品特定の可能性を検証		✓ 画像や動画を含むデータ

Copyright © 2020 Accenture All rights reserved

品番・メーカーフラグの作成

- ✓ 品番別・メーカー別に分析を進めるために、品番一覧およびメーカー一覧を作成
- ✓ 作成した品番一覧・メーカー一覧をマスタとし、取得済のSNSデータに対して、品番フラグおよびメーカーフラグを立てて後続の分析を実施

品番・メーカー一覧の作成方法

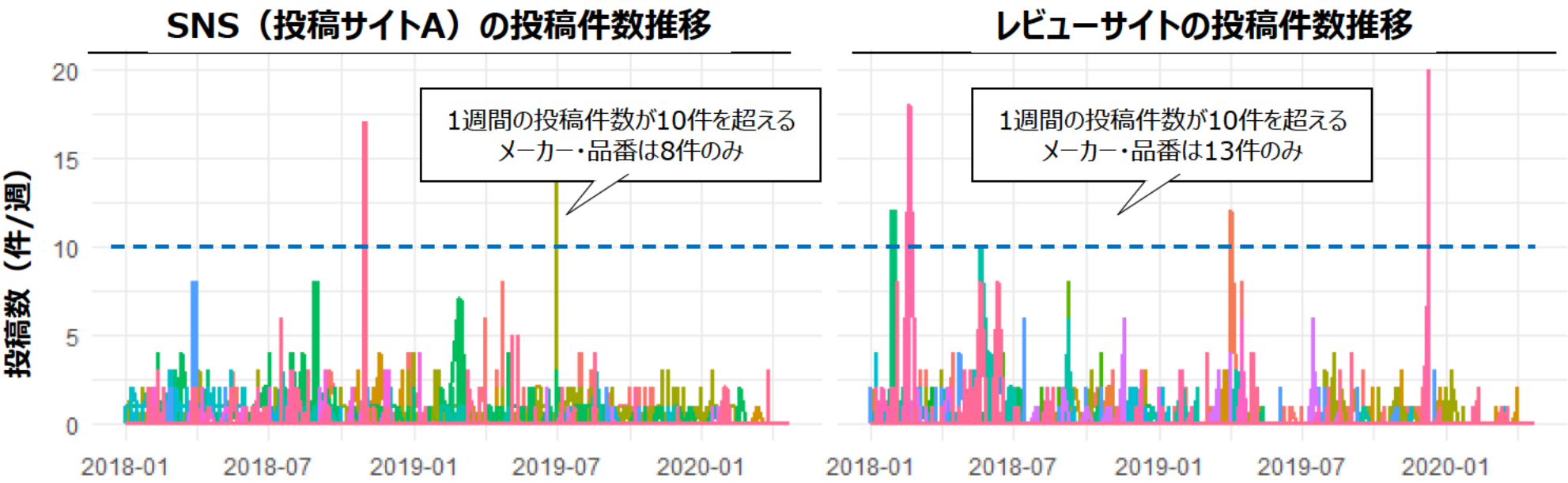
品番	**「先頭の文字が英字で数字を含む3文字以上の文字列」を品番と定義**してSNSの投稿データから品番文字列を抽出
メーカー	**過去の重大事故データベースおよびレビューサイト中のメーカー一覧から、対象品目と関連するメーカー名を抽出**

→ 作成した品番・メーカー一覧とSNSの投稿データとをマッチングし、**品番・メーカーフラグを立てる**

Copyright © 2020 Accenture All rights reserved

SNSデータの週別の投稿数推移

✓ メーカー名または品番を含むSNS投稿を週別に集計した結果、週あたりの投稿数が10件を超えるものは2018年以降21件（レビューサイト:13件他）のみであり、大部分は10件以下と少数の投稿のみであった

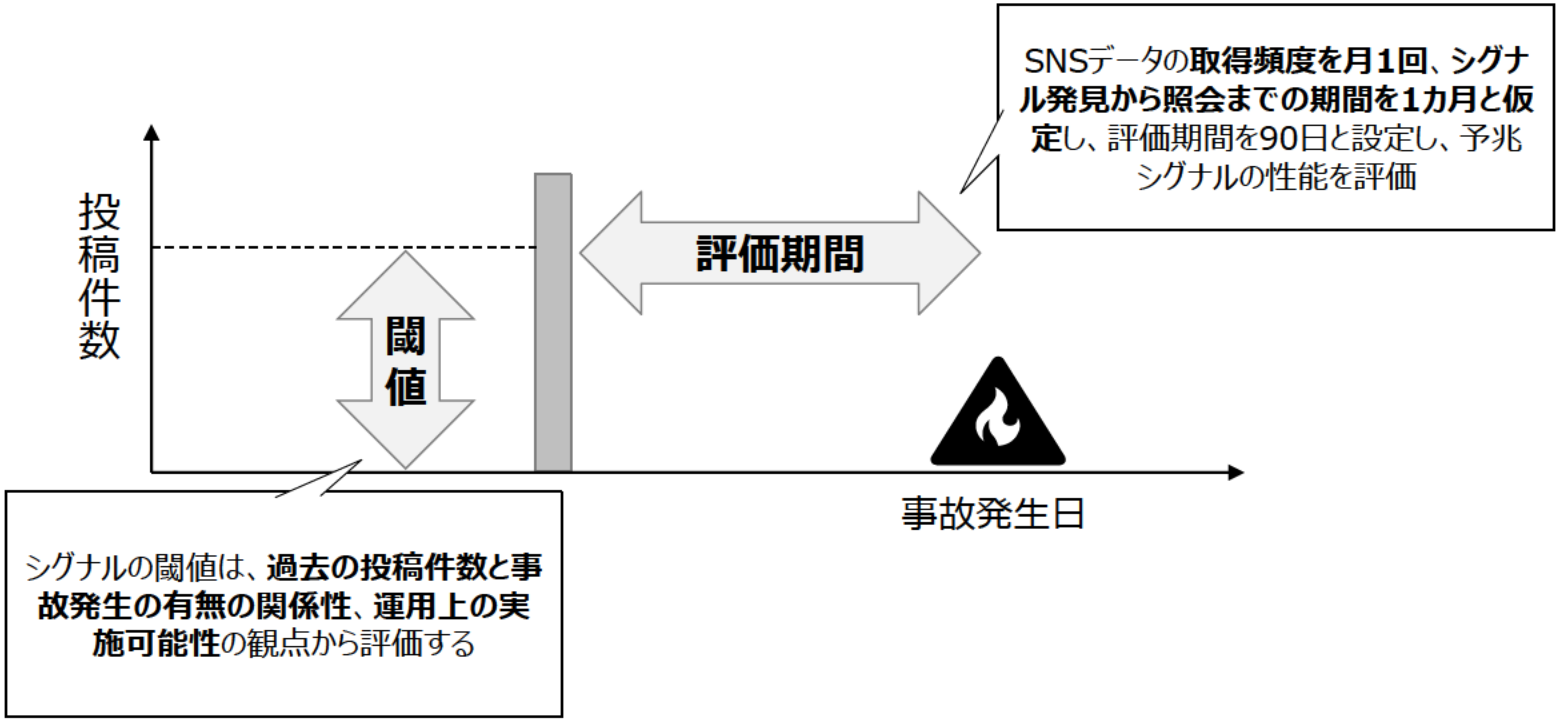


※メーカー名・品番を含む投稿を当該メーカー・品番ごとにプロットして重ね合わせたグラフ

Copyright © 2020 Accenture All rights reserved

予兆シグナルの性能評価：定義

✓ 事故予兆の指標（シグナル）は、「**投稿件数**（週別）」「**評価期間**（投稿から何日後までの事故を評価対象とするかの期間）」の観点から定義する

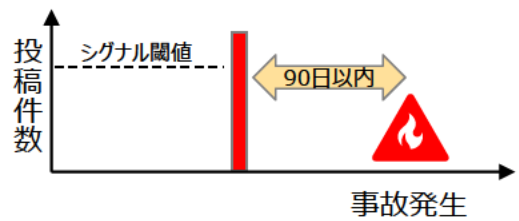


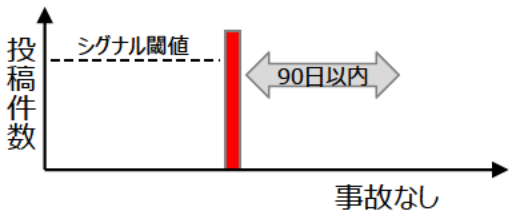
Copyright © 2020 Accenture All rights reserved

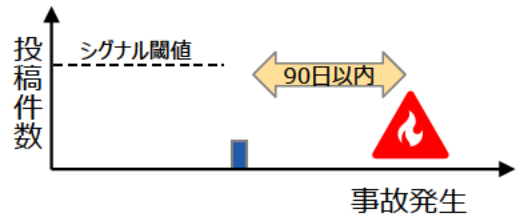
予兆シグナルの性能評価：評価方法

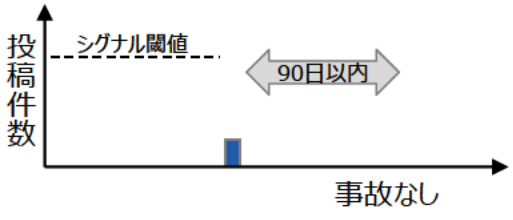
- ✓ 投稿件数に対して複数の閾値を設定し、当該シグナル発生から90日以内の事故有無との関係を評価
- ✓ 投稿件数が閾値以上（シグナルあり）か未満（シグナルなし）に分類し、2x2の混同行列を作成し性能評価

		90日以内に事故発生あり	90日以内に事故発生なし
投稿件数が閾値以上	シグナルあり	**真陽性**	**偽陽性**
投稿件数が閾値未満	シグナルなし	**偽陰性**	**真陰性**









Copyright © 2020 Accenture All rights reserved

予兆シグナルの性能評価：評価指標

✓ 混同行列を用い、シグナルの有無と実際の事故発生有無を比較することで、シグナルの予兆性能を評価する

混同行列

件数	事故発生 あり	事故発生 なし
シグナル あり	①真陽性 (TP:True Positive)	②偽陽性 (FP:False Positive)
シグナル なし	③偽陰性 (FN:False Negative)	④真陰性 (TN:True Negative)

混同行列を用いた指標

正解率 (Accuracy)	$\frac{TP + TN}{TP + FP + FN + TN}$	• 全ての判定対象のうち、正しく予測された件数
誤分類率 (Misclassification rate)	$\frac{FP + FN}{TP + FP + FN + TN}$	• 全ての判定対象のうち、誤って予測された件数
適合率 (Precision)	$\frac{TP}{TP + FP}$	• シグナルありと判定されたもののうち、正しく予測された件数
再現率 (Recall)	$\frac{TP}{TP + FN}$	• 実際に事故ありだったもののうち、正しく事故ありと予測できた件数
F値	$2 \times \frac{適合率 \times 再現率}{適合率 + 再現率}$	• 適合率と再現率の調和平均で両者のバランスをとった指標

使用分野や予測目的によって重要視される指標は異なる（以下例）

- クレジットカードの不正使用のように、一つの不正が大きな損失につながる場合・・再現率を重要
- 開封確率や購入確率の高い層に向けて効果的にDM送付をしたい場合・・・適合率を重視

※各指標とも0～1の範囲の値を取る
※※誤分類率を除き、値が高いほど良い指標

Copyright © 2020 Accenture All rights reserved

予兆シグナルの性能評価：評価結果

- ✓ 過去の事故発生と投稿（重大事故ワードを含む）の関係性を評価した結果、週あたりの投稿件数が1件以上をシグナルと定義した場合に、事故予兆の性能が最も高くなった（適合率と再現率のバランスが最も良い）
- ✓ ただし、品目によって適合率・再現率は異なる

重大事故ワード（火、人への傷害）に関する投稿の予兆シグナル性能

ソース	品目	最適閾値	混同行列				正解率 Accuracy	適合率 Precision	再現率 Recall	F値
			TP	FP	FN	TN				
SNS (投稿サイトA)	ノートパソコン	1	60	103	525	789	57.5%	36.8%	10.3%	16.0%
	バッテリー	1	121	106	246	711	70.3%	53.3%	33.0%	40.7%
	携帯電話機	1	117	58	353	625	64.4%	66.9%	24.9%	36.3%
	照明器具	1	4	11	219	363	61.5%	26.7%	1.8%	3.4%
	電気ストーブ	1	1	8	110	460	79.6%	11.1%	0.9%	1.7%
レビューサイト	ノートパソコン	1	20	20	445	684	60.2%	50.0%	4.3%	7.9%
	バッテリー	1	14	26	337	873	69.8%	35.0%	3.8%	6.8%
	携帯電話機	1	9	7	351	402	53.4%	56.3%	2.5%	4.8%
	照明器具	1	6	11	146	517	76.9%	35.3%	3.9%	7.1%
	電気ストーブ	1	9	15	106	392	76.8%	37.5%	7.8%	12.9%

※最適閾値：網羅的に投稿件数の閾値を変化させたときに予兆性能が一番良い（F値が一番高い）閾値を選択
※※重大事故ワードを含みかつ、メーカー名又は品番を含む投稿の出現頻度は低いこともあり、最適閾値は1件となっている（次頁参照）

Copyright © 2020 Accenture All rights reserved

メーカー名・品番を含むSNSデータの割合

✓ メーカー名又は品番を含み、かつ重大事故ワード（火、人への傷害）を含む投稿は全体の3～5%程度であり、出現割合は低い

投稿件数(%)

		事故関連ワード・品目を含む投稿件数	メーカー名又は品番を含む投稿	重大事故ワード含む*
全体		254,248	42,647 (16.8%)	12,517 (4.9%)
ソース	SNS(投稿サイトA)	173,676	28,365 (16.3%)	10,151 (5.8%)
ソース	レビューサイト	80,572	14,282 (17.7%)	2,366 (2.9%)

評価期間：2016年12月～2020年11月
＊：「火」、「人への傷害」に関する関連ワードを含む投稿

Copyright © 2020 Accenture All rights reserved

予兆シグナルの閾値：運用上確認可能な投稿件数の見積もり

✓ 週1回半日程度の作業時間を確保できれば、シグナル閾値を1件とした場合でも十分対処可能

週あたりの処理可能な投稿件数

＜前提条件＞

作業時間：5時間/週（週1回半日程度）

事前準備：1時間

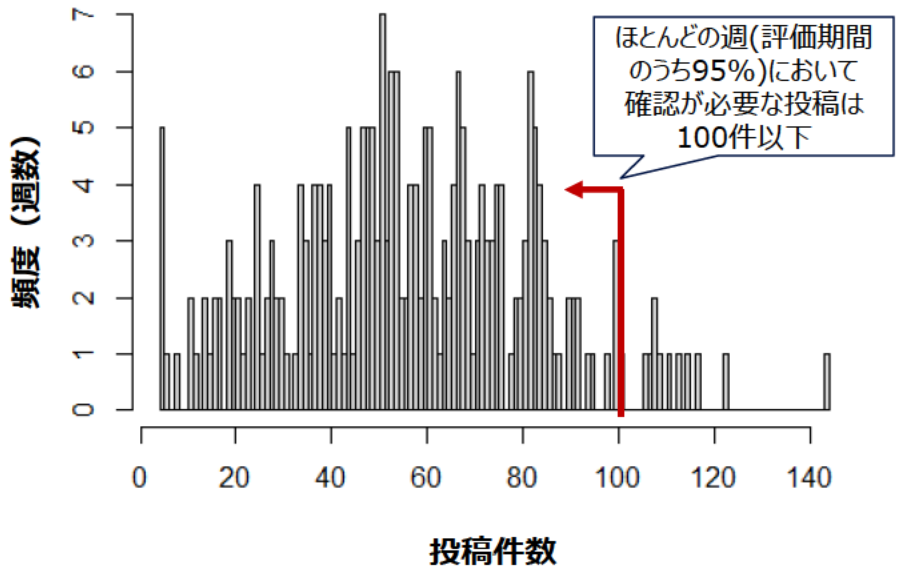
投稿内容の目視確認：30秒/件

詳細深堀り（他ソースからの情報収集等）：30分/件

詳細深堀りが必要な投稿割合：1%（次頁参照）

1週間で処理可能な投稿件数は300件
（詳細深堀りが必要な投稿は3件）

確認が必要な投稿件数のヒストグラム



※シグナル閾値を1件とし、重大事故ワード（火、人への傷害）、メーカー名・品番を含む投稿の週あたりの合計件数（2016年11月～2020年11月）

閾値を1件とした場合でも想定作業時間を確保できれば、十分対応可能と考える

Copyright © 2020 Accenture All rights reserved

ノイズ投稿割合の目視確認

- ✓ 2020年1月から11月における重大事故ワード（火、人への傷害）を含み、メーカー名又は品番を含む投稿について、事故との関連性のある投稿数を目視にて確認
- ✓ 関連性のある投稿の割合は1%未満であり、ほとんどの投稿は事故との関連性のないノイズ投稿であった

SNSデータ

検索条件
期間：2020年1月～11月
事故関連ワード：火、人への傷害
品番またはメーカー名を含む投稿

→ 目視確認で事故との関連性を評価

事故との関連性の割合評価

区分		確認件数	関連性あり(%)
全体		2,056	8 (0.4%)
ソース	SNS(投稿サイトA)	1,825	6 (0.3%)
ソース	レビューサイト	231	2 (0.9%)
フラグ	メーカー名含む	1,261	3 (0.2%)
フラグ	品番含む	795	2 (0.6%)

- **SNS(投稿サイトA)よりもレビューサイトの方が関連割合は高く、メーカー名を含む投稿よりも品番を含む投稿の方が事故関連割合は高い傾向**
- いずれも1%未満であることから、**ノイズが非常に多い**

Copyright © 2020 Accenture All rights reserved

シグナルの閾値　まとめ

シグナルの予兆性能	運用上の実施可能性	SNS特有のノイズ
過去の事故発生パターンと投稿数の関係性から、1件以上をシグナルとすることで、事故予兆の性能が最も高かった	週半日程度の稼働でも、メーカー名又は品番を含む投稿の全件チェックが可能	事故関連ワードを含み、メーカー名又は品番を含む投稿であっても事故との関連性が示唆されるものは1%程度と少ない

閾値を1件とし、事故関連ワードおよびメーカー名・品番を含む投稿については、
元データの投稿内容を確認するのが良いと考える

Copyright © 2020 Accenture All rights reserved

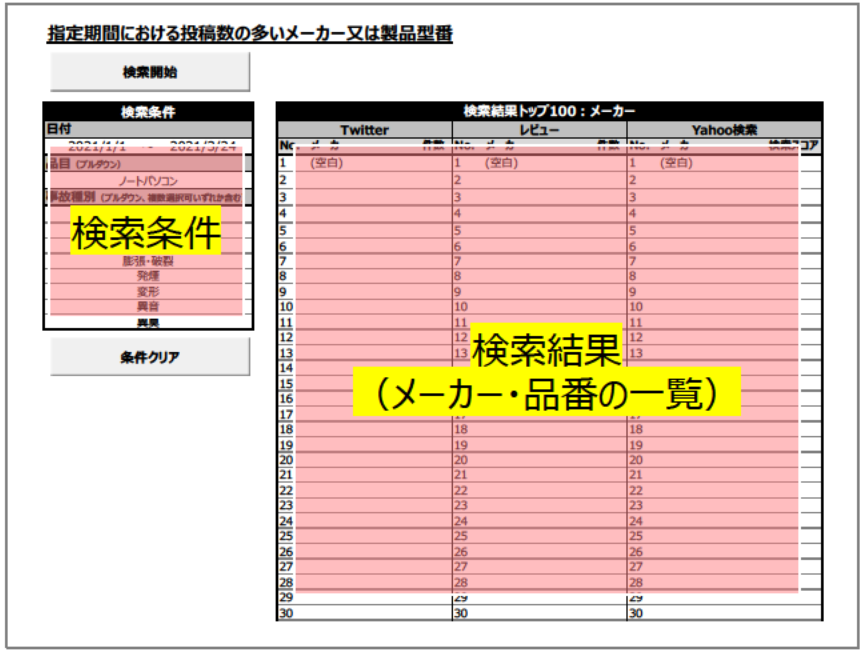
3. 予兆分析の成果　3-5. ダッシュボード作成

簡易ダッシュボードの概要

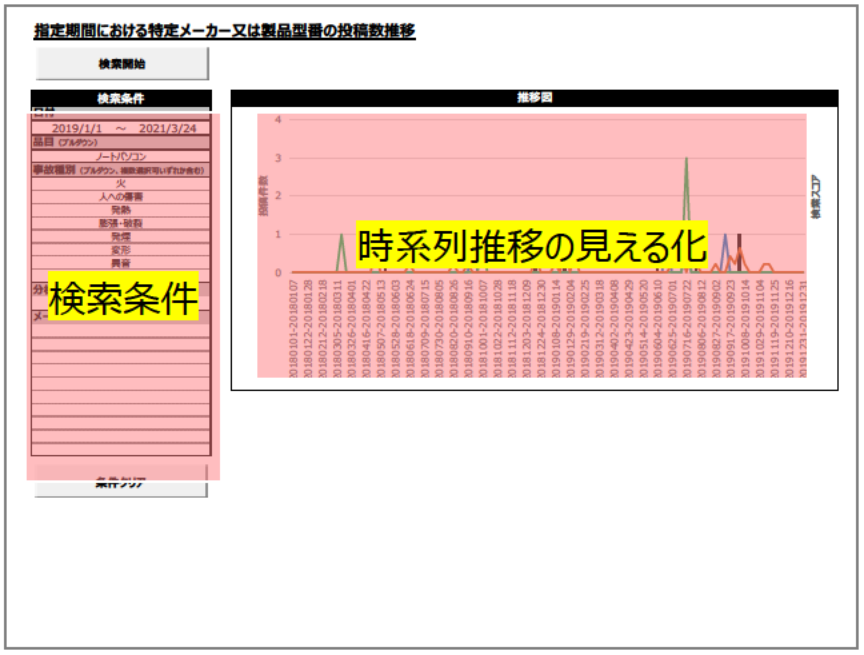
- ✓ メーカー名および品番を含むSNS投稿から事故の予兆検知／再発防止につなげる
- ✓ シグナル検知のためのCSVファイルを自動生成し、簡易ダッシュボード（Excel）に読み込ませることで更新可能
- ✓ メーカー名又は製品型番等を選択することで必要情報の閲覧が可能（迅速に確認可能）
- ✓ 消費者の関心の推移を把握し、事故予兆を探る仮説立案の材料とする（新たな品番の発見等）

簡易ダッシュボード画面例

①メーカー名又は品番の特定



②シグナル検知



Copyright © 2020 Accenture All rights reserved

本レポートのサマリとの対応（予兆分析）

インターネット検索エンジンにおける検索データおよびSNSデータ等を収集し、製品事故の予兆を捉えることができるかを分析・検証を実施した。その結果を踏まえ、予兆を捉えるためのツールイメージを作成

□予兆分析：

- ✓ 検索データを分析し、「品番」×「事故ワード」等の検索量により予兆検知の可能性を複数発見
- ✓ SNSデータと検索データを含め予兆検知の可能性を分析し、「品番又はメーカー名が含まれる投稿」をシグナルとして用いる予兆検知方法を検証
- ✓ 上記を実現するツールイメージを作成

□今後に向けた課題：

- ✓ 品番一覧・メーカー名一覧の拡充
- ✓ SNSデータのクリーニング精度向上
- ✓ 継続したデータ収集と運用ノウハウの拡充

Copyright © 2020 Accenture All rights reserved

3. 予兆分析の成果

まとめ

	サマリ	実施項目
（1）法令違反製品・販売事業者の傾向分析	Webクローリングによるデータ収集の技術的有用性を示すとともにデータクレンジングについての運用示唆をまとめ、15個の仮説検証および機械学習モデル構築・精度検証を実施し、実運用に向けた示唆を提示した。	✓ データ収集・データクレンジング完了 ✓ 法令違反確認完了 ✓ 15仮説の検証完了 ✓ 機械学習モデル構築・精度検証完了 ✓ ネトパト事業連携実施完了
（2）重大・非重大製品事故の予兆分析	検索データによる予兆検知の可能性を複数発見、SNSデータを含めて予兆検知の可能性を分析し、「品番又はメーカー名が含まれる投稿」をシグナルとして用いる予兆検知方法を提示した。 上記を実現するツールイメージも作成した。	✓ 検索データ収集・検証完了 ✓ SNSデータ収集完了 ✓ SNSデータ分析完了 ✓ ダッシュボードツールイメージ作成完了
（3）分析結果の妥当性検証	第三者有識者３名にヒアリングを実施し分析アプローチ等の妥当性を確認した。	✓ 有識者３名調整・案件説明完了 ✓ レポート提示・ヒアリング完了

Copyright © 2020 Accenture All rights reserved

4.運用に向けた考察

Copyright © 2020 Accenture All rights reserved

目次

1. 調査の概要
- 1-1. 調査の背景・目的
- 1-2. スケジュール
- 1-3. サマリ

2. 傾向分析の成果
- 2-1. 作業の全体像
- 2-2. クローリング技術検証
- 2-3. データ収集
- 2-4. データクレンジング
- 2-5. 法令違反等の確認
- 2-6. データ分布状況の確認
- 2-7. 仮説選定・検定
- 2-8. 機械学習モデル構築・精度検証

3. 予兆分析の成果
- 3-1. 作業の全体像
- 3-2. 検索データ収集・分析結果
- 3-3. SNSデータ収集・分析結果
- 3-4. 予兆検知手法検証
- 3-5. ダッシュボード作成

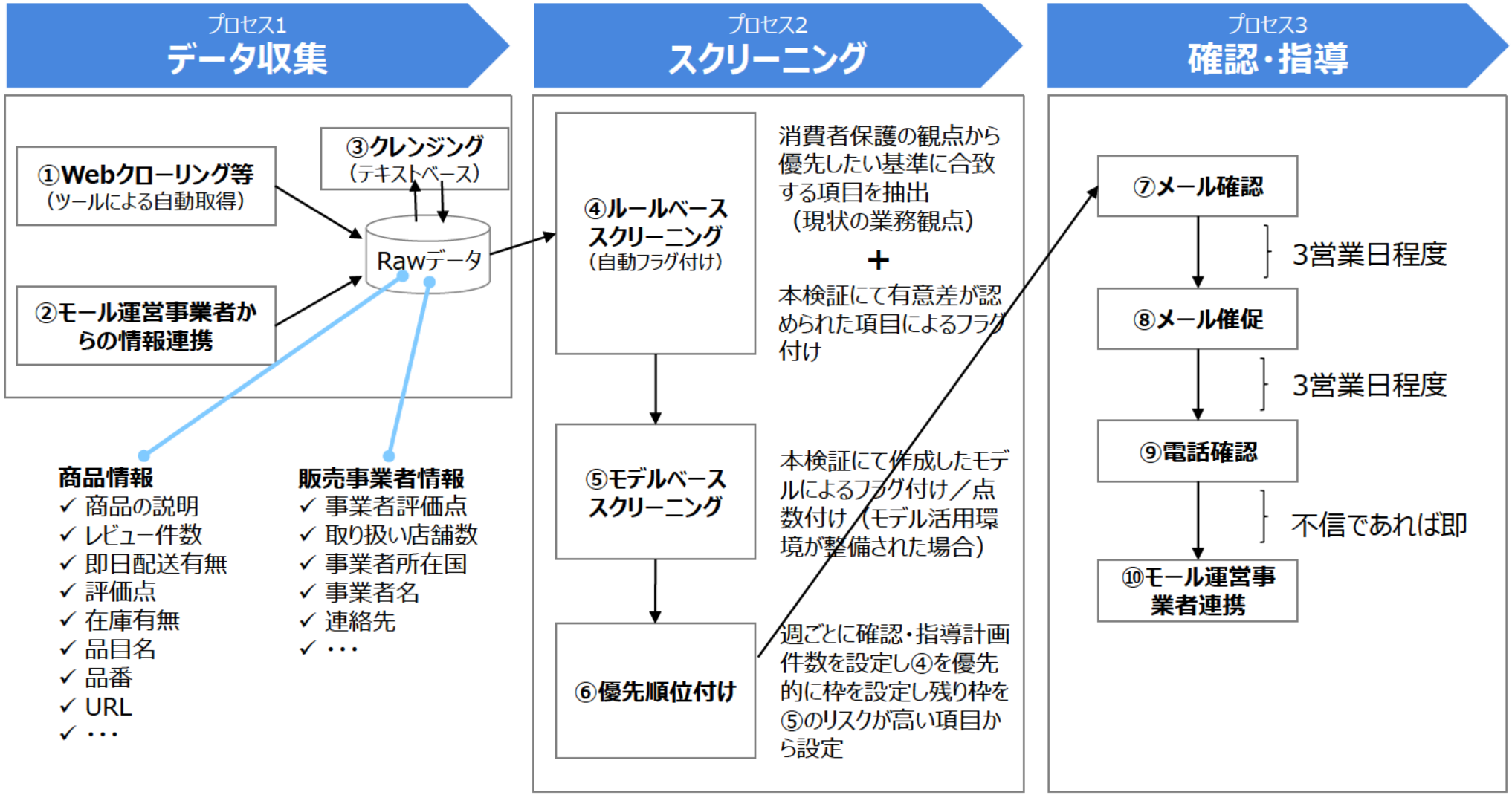
4. 運用に向けた考察
- 4-1. 傾向分析の活用とモデル運用
- 4-2. 予兆分析ダッシュボードの運用
- 4-2. ネットパトロール効率化

5. 運用に向けた課題・対応策
6. 第三者有識者コメント

Copyright © 2020 Accenture All rights reserved

ネットパトロール事業運用改善提案

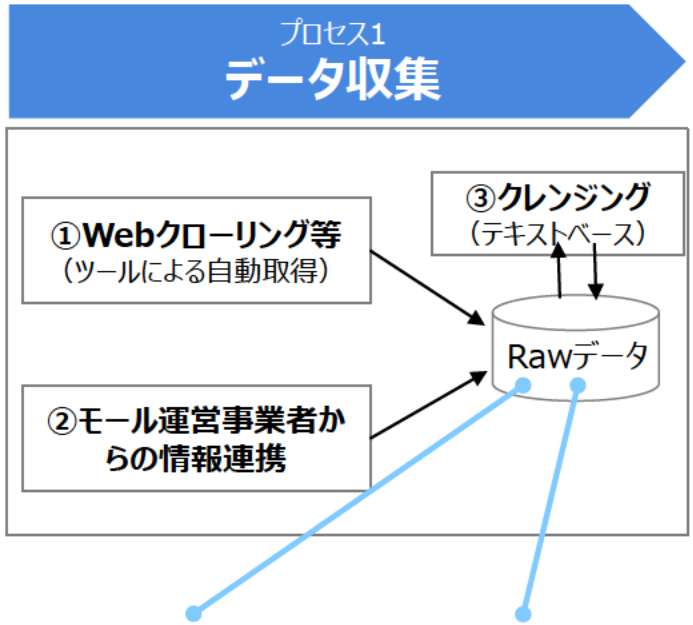
本年度の調査結果を活用した効率的・効果的なスクリーニングを提案



Copyright © 2020 Accenture All rights reserved

ネットパトロール事業運用改善提案

Webクローリング等により網羅的かつタイムリーなデータ収集を実現



商品情報
- ✓ 商品の説明
- ✓ レビュー件数
- ✓ 即日配送有無
- ✓ 評価点
- ✓ 在庫有無
- ✓ 品目名
- ✓ 品番
- ✓ URL
- ✓ ・・・

販売事業者情報
- ✓ 事業者評価点
- ✓ 取り扱い店舗数
- ✓ 事業者所在国
- ✓ 事業者名
- ✓ 連絡先
- ✓ ・・・

➢ 課題意識

現状のデータ収集方法はマニュアル運用での取得であるため
ECサイト上の情報を**網羅的かつタイムリー**に確認することが困難
属人的に発見した疑義商品を確認・指導しているのが現状と認識

➢ 解決策

Webクローリングツールによる２週間に１回程度（５品目程度）
またはモール運営事業者からの情報提供依頼が有効

➢ 検証結果からの示唆

①の作業を検証機関に実施したところ
Webクローリングによる収集が可能であることを確認した
ノイズ混入（対象外商品の混入）は不可避であり
③について作業者１名～２名程度の稼働により
テキストベース（キーワード抽出）によるクレンジングが可能であった

➢ 発展提案

モール運営事業者から情報を定期的に提供いただくことで
より正確な情報をタイムリーに入手できるのではないか
今回の仮説検証により有意であった項目には
秘匿性が低いと考えられる項目であるため
実現は難しくないと考えられる

Copyright © 2020 Accenture All rights reserved

ネットパトロール事業運用改善提案

本年度の仮説検定結果とモデルを活用し効果的なスクリーニングを実現

プロセス2
スクリーニング

④ルールベース スクリーニング
（自動フラグ付け）

消費者保護の観点から優先したい基準に合致する項目を抽出
（現状の業務観点）
＋
本検証にて有意差が認められた項目によるフラグ付け

↓

⑤モデルベース スクリーニング

本検証にて作成したモデルによるフラグ付け／点数付け

↓

⑥優先順位付け

週ごとに確認・指導計画件数を設定し④を優先的に枠を設定し残り枠を⑤のリスクが高い項目から設定

➢ 課題意識

熟練者の経験等に基づく検索キーワードにより対象の選定を行っている
選定基準が統計的に有意であるか確認できていないのではないか
（傾向の変化への追従や脱属人化という課題もある）

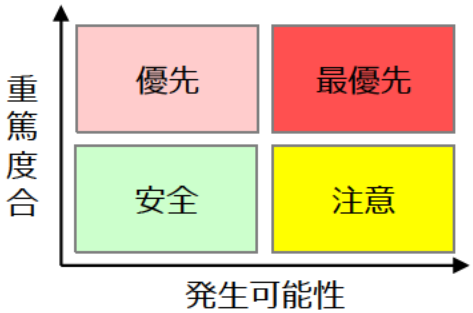
➢ 解決策

熟練者の基準に加え、統計的優位性が示された基準と
モデルによる予測結果を組み合わせて活用することで、
選定基準を明確化し優先順位を戦略的に設定可能とする

➢ 発展提案

週に20件確認するなど計画値を設定し優先度に応じて
ランダムサンプリング率を調整する
④および⑤のフラグ付けを発生可能性ととらえ
重篤度と合わせて以下のように優先順位付けする

重篤度を
品目により
点数化



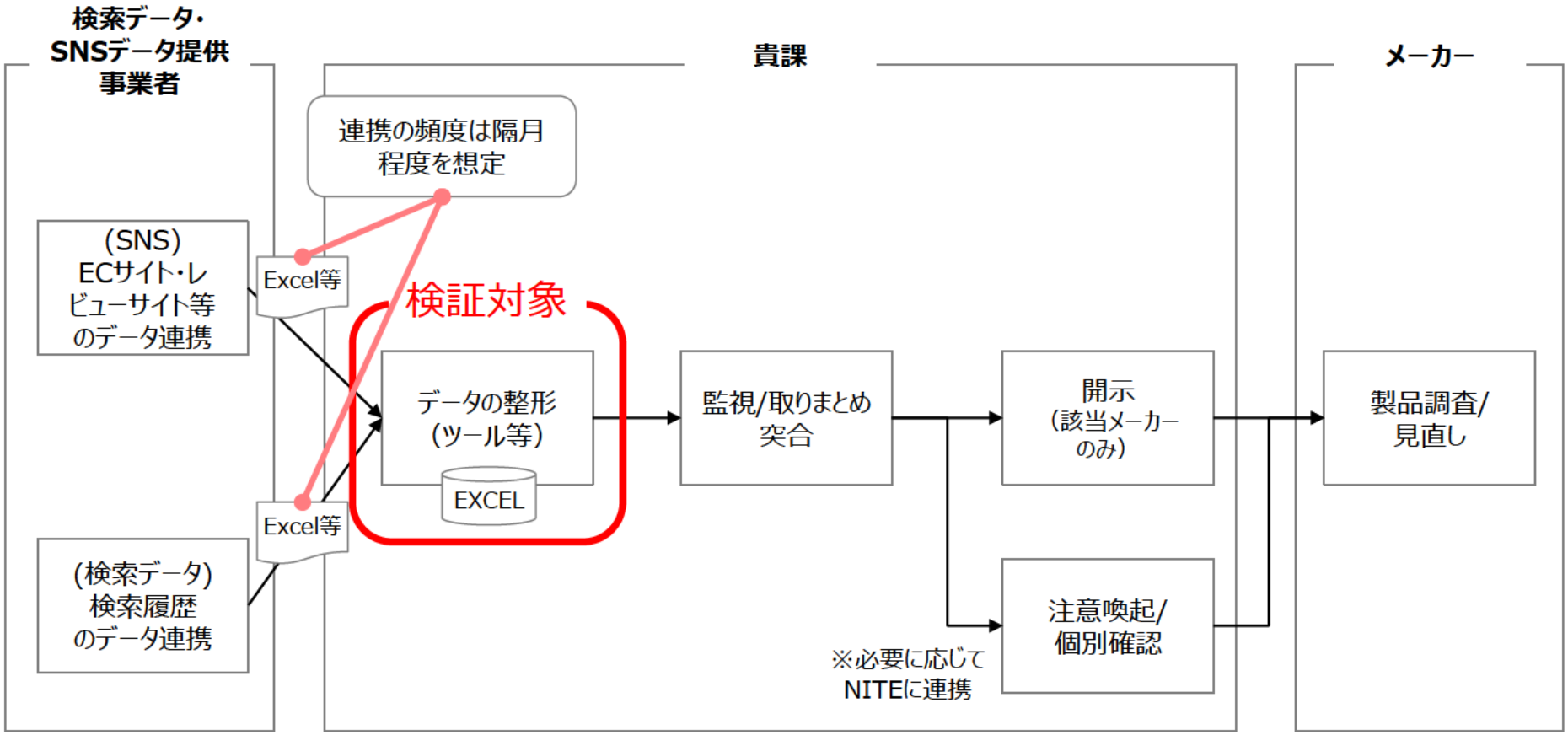
ランダムサンプリング率を
優先度により調整
（安全でも０としない）

Copyright © 2020 Accenture All rights reserved

予兆分析の運用に向けた考察

- ✓ 品番による検索が多いバッテリーや携帯電話機を定期監視することで予兆検知の可能性あり
- ✓ 電気ストーブのように品番検索が少ないがリスクが高い製品を定期監視に含めることで予兆検知の可能性はある
- ✓ 品番×事故ワードによる検索の監視が有効な商品を定期監視しながら、リスクの高い商品もあわせて監視する運用が想定される

（将来像）分析結果の活用イメージ



Copyright © 2020 Accenture All rights reserved

５.運用に向けた課題・対応策

Copyright © 2020 Accenture All rights reserved

目次

1. 調査の概要
- 1-1. 調査の背景・目的
- 1-2. スケジュール
- 1-3. サマリ

2. 傾向分析の成果
- 2-1. 作業の全体像
- 2-2. クローリング技術検証
- 2-3. データ収集
- 2-4. データクレンジング
- 2-5. 法令違反等の確認
- 2-6. データ分布状況の確認
- 2-7. 仮説選定・検定
- 2-8. 機械学習モデル構築・精度検証

3. 予兆分析の成果
- 3-1. 作業の全体像
- 3-2. 検索データ収集・分析結果
- 3-3. SNSデータ収集・分析結果
- 3-4. 予兆検知手法検証
- 3-5. ダッシュボード作成

4. 運用に向けた考察
- 4-1. 傾向分析の活用とモデル運用
- 4-2. 予兆分析ダッシュボードの運用
- 4-2. ネットパトロール効率化

5. 運用に向けた課題・対応策

6. 第三者有識者コメント

Copyright © 2020 Accenture All rights reserved

5. 運用に向けた課題・対応策

残課題・対応策（傾向分析）

✓ 法令違反サンプルの効率的な収集および継続的な分析・検証等を実現していくことが今後の課題

No.	課題	対応策
1	法令違反サンプル不足	✓ **法令違反サンプル収集の仕組みづくり**（例えばモール運営事業者と連携の上、PSマーク画像掲載をルール化することで、法令違反照会の手間を大幅に減らす仕組みづくり） ✓ 教師なし学習による補完（異常検知*等）
2	対象品目の偏り	✓ 他品目の法令違反サンプルの収集・検証（検証をスケールさせ、より幅広い品目で仮説検証を進めることで更なる業務効率向上を図る） ✓ 品目に依存しない品目横断的な特徴量の追加検討
3	継続的なデータ収集・分析・検証	✓ 継続的にデータ収集・分析・検証を行うことで最新のトレンドへのキャッチアップと分析精度向上を図る
4	予測モデルの更なる進化	✓ 品目横断的モデルと各品目の特性を捉えた個別モデルとを最適に組み合わせたモデルの発展形（アンサンブルモデル等）

＊：法令違反サンプルの取得が困難であることから、法令違反ラベルを使わない分析手法（例えば異常検知アルゴリズム）を用いて、他の販売事業者と異なる特徴・挙動を示す販売事業者を特定し、その傾向を評価することで新たな仮説の創出および高度化につなげていくことが可能

Copyright © 2020 Accenture All rights reserved

残課題・対応策（予兆分析）

✓ 品番一覧、メーカー名一覧の拡充とSNSデータのクリーニング精度向上、運用ノウハウの拡充が今後の課題

No.	課題	対応策
1	品番一覧・メーカー名一覧の拡充	✓ 情報ソースを広げ、品番およびメーカー名を拡充し網羅性向上と最新情報へのキャッチアップを図る
2	SNSデータのクリーニング精度向上	✓ 運用しながらSNSデータのノイズを除去するための除外ワード等の拡充により継続的に精度改善
3	継続したデータ収集と運用ノウハウの拡充	✓ データ収集（データの購入等による）を実際に行う中で、最適な収集頻度等のノウハウを蓄積しより効率的な業務の実現を図る

Copyright © 2020 Accenture All rights reserved

6. 第三者有識者コメント

Copyright © 2020 Accenture All rights reserved

目次

1. 調査の概要
- 1-1. 調査の背景・目的
- 1-2. スケジュール
- 1-3. サマリ

2. 傾向分析の成果
- 2-1. 作業の全体像
- 2-2. クローリング技術検証
- 2-3. データ収集
- 2-4. データクレンジング
- 2-5. 法令違反等の確認
- 2-6. データ分布状況の確認
- 2-7. 仮説選定・検定
- 2-8. 機械学習モデル構築・精度検証

3. 予兆分析の成果
- 3-1. 作業の全体像
- 3-2. 検索データ収集・分析結果
- 3-3. SNSデータ収集・分析結果
- 3-4. 予兆検知手法検証
- 3-5. ダッシュボード作成

4. 運用に向けた考察
- 4-1. 傾向分析の活用とモデル運用
- 4-2. 予兆分析ダッシュボードの運用
- 4-2. ネットパトロール効率化

5. 運用に向けた課題・対応策

6. 第三者有識者コメント

Copyright © 2020 Accenture All rights reserved

6. 第三者有識者コメント

有識者ヒアリング サマリ

✓ 分析アプローチ・解釈の仕方について妥当であることを確認いただいた。また、複数の改善案をコメントいただいたので、今後の改善案として整理

	統計有識者	機械学習有識者	実務有識者	改善案
仮説検証	• 複数の特徴量を同時に評価できる**教師なし学習**（潜在クラス分析等）を用いて製品を類型化すると解釈しやすい • **複数要因の連関を考慮**した仮説構築するとよい		• データが少ない件については**モール間のクロスチェック**（ブラックリストを作りモール横断で調査）を検討するとよい	• 要因を複数組み合わせた仮説検証の実施 • 違反傾向の高い事業者については、その情報を別モールのデータにもあてはめて評価
予測モデル		• 特徴量同士の相関が高い項目は絞り込むとよい • 品目別モデルはデータ量が少なく、データ拡充、**不均衡データへの対応**などが必要	• 何をもって有用と判断するかの基準を明確にしておくとよい	• 不均衡（違反例が少ない）の対応
総合	• 様々な検知技法があるため、**常に改良を意識した取り組み**をすることが好ましい	• 実施した検証結果を活用するにあたり、モデルの再現率や適合率などの特徴を把握しておくことが望ましい	• 実際に起きた違反事例の特徴を踏まえて**特徴量を補強**すると良い • **継続的なデータ収集＆モデル更新**が必要	• 違反事例の特徴を踏まえた特徴量の補強 • 継続的なデータ収集・モデル更新の仕組み作り

Copyright © 2020 Accenture All rights reserved