

GROWTHY, Essential Platform for Data-driven

#big data #data science #DataOps



JaeYong Choi, HyeongBae Shin, Taeho Yun

LINE / 開発2センター ビッグデータプラットフォーム室、
LINE+/Game Platform Engineering/Big Data Platform、
LINE / 開発2センター ビッグデータプラットフォーム室

If you can not measure, you can not manage.

Data-driven - Making decisions based on Data

Data-drivenとは、様々なBig Dataを活用する企業で個人の経験、直感等を通じて事業を展開することなく、蓄積してきたDataを基にビジネスの方向を確立し、実行していくことです。LINEはLINE Styleという6つのキーワードを通じて会社の文化を作っており、その中にこのData-drivenという概念も含まれています。

LINE Style: NEEDS / SPEED / DETAIL / TEAMWORK / ENJOY / DATAの内、DATAとは、DATAに基づいたビジネス活動することを意味しており、実際にそういった形で業務を進めています。なお、Data-drivenを通じて成功的なビジネス活動をするためには、基礎として「2つのData関連部分」と「3つの分析部分」という要素が必須であり、それぞれはGROWTHYというプラットフォームに含まれています。

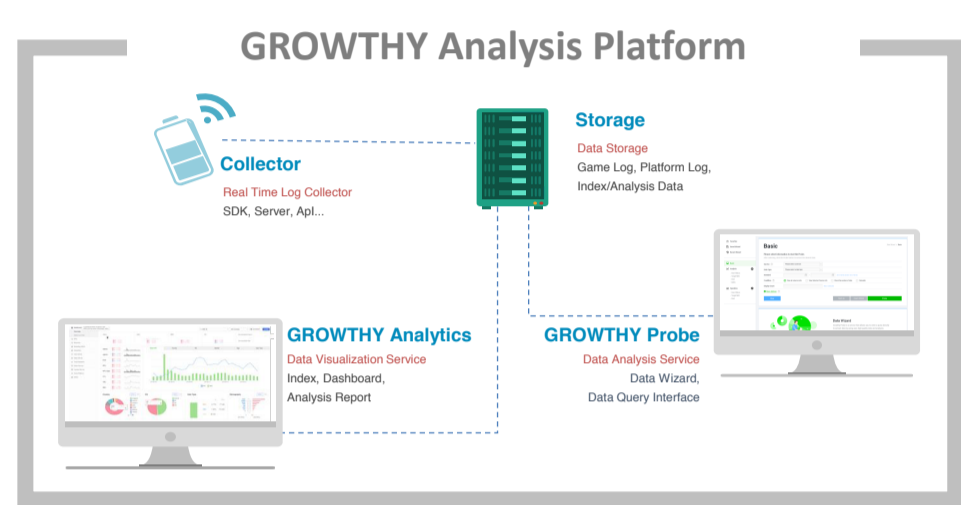
【必須の機能】

- データの基準の統一性(Uniformity of Data standards): 同じ事業群のデータは、基準が同一でなければなりません。
- データの正確性(Data Quality): データがベースとなるため、基礎品質に常に問題がないようにしなければなりません。
- 状況分析(Standard Reports and AdHoc Reports Analysis): 意思決定のための迅速な状況の分析が必要です。
- リスク分析(Risk Management Reports Analysis): リスク分析は、突然の危険を未然に防ぐために必要です。
- 予測分析(Generic Predictive Analysis): 過去のデータを通じて、将来成功するかしないかを予測することができます。

上記のように基礎データが安定していることで、状況分析から予測分析までDataを通じたDecision Makingが可能になります。

私たちは、そういったData-drivenな意思決定の可能なPlatformを創るために多くの開発者やアナリストが一つになって様々な工夫をし、多くのコンテンツを作成してきました。データ関連業務に携わっている、あるいは勉強中の方々も一度は同じ悩みをしたことがあるのではないかと思います。素晴らしいOrchestraのようなプラットフォームを作るために始めた様々なことが、本日少しでもお伝えできればと思います。

それでは、本格的なData-drivenのための統合されたLINEのGROWTHY Analysis Platformをご紹介します。さらにその中のコンテンツと機能についても説明します。

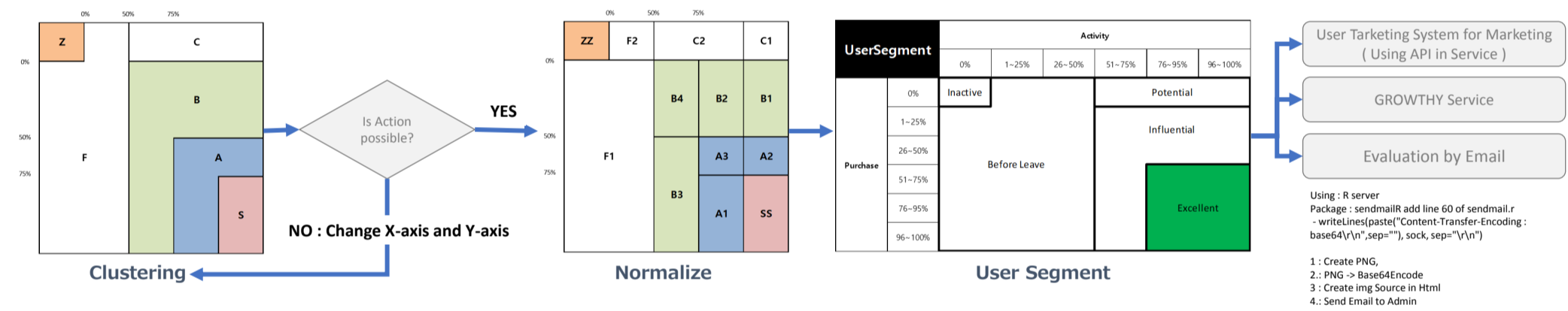


- > Collector: SDKを通じてデータを収集
- > Storage: Hadoop Eco System基盤のデータ保存
- > GROWTHY Analytics: 伝統的なBIシステムと類似性はあるものの、Pivot基盤の定型サービスを提供
- > GROWTHY Probe: Raw dataを通じてSQL基盤のSelf分析環境を提供

We support a lot of contents for Data-driven through GROWTHY Analytics and Probe

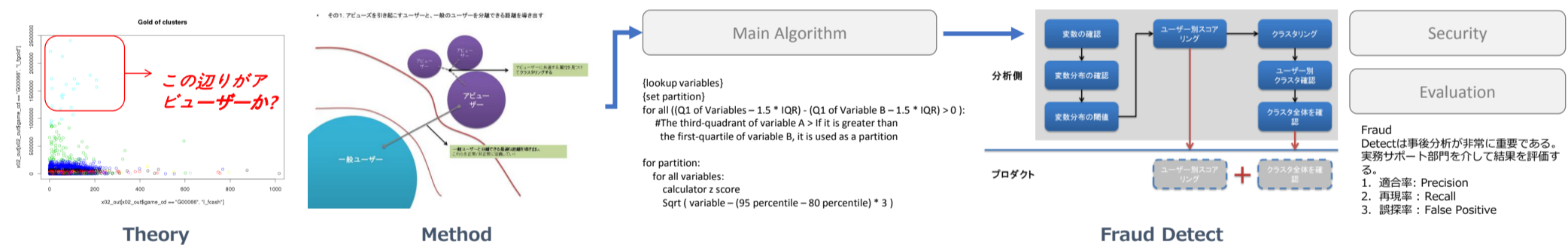
User Segment Analysis

Service状況の分析をするとき、基本的なDataでEDA (Exploratory Data Analysis) をしたり、重要なKPIのTrends把握を通じて、分析を始めることになります。しかし、そのような分析はServiceの変化スピードに追いつけず、失敗になることが多いです。数回は可能ですが、無数の問題を解決していくには、重要度に応じた解決策を考え出す必要があります。そのような部分を私たちは、User Segmentを通じて事前に分類し、Business的な理解を通じてアクションが可能な結果を作りました。つまり、ユーザーの活動性と課金状況により、5つのグループに分けることが可能です。こういったユーザーの細分化は、データを基にすることで目的に合わせて追加の拡張が可能となり、様々なマーケティング、ユーザー分析、原因分析などに活用可能になります。



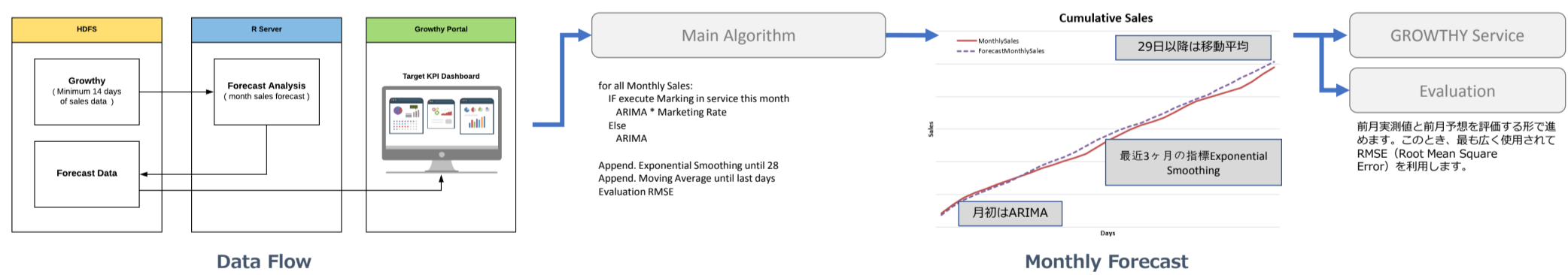
Frauds Detect Analysis

Fraud Detect Analysisは、Risk Managementのために必要で、異常な行動がデータを通じて把握された場合、対応することが可能です。様々な異常行為は、データを通じて、異常値の検出をしたり、そのパターンを見つけるなどの方法的な部分を保有し、教師なし (Unsupervised) の形で分析することになります。教師あり (Supervised) 学習が可能であれば分析手法がさらに多様化しますが、判別されていないユーザーを通じて選抜をする必要があるため、まず教師なし (Unsupervised) の形で進めます。



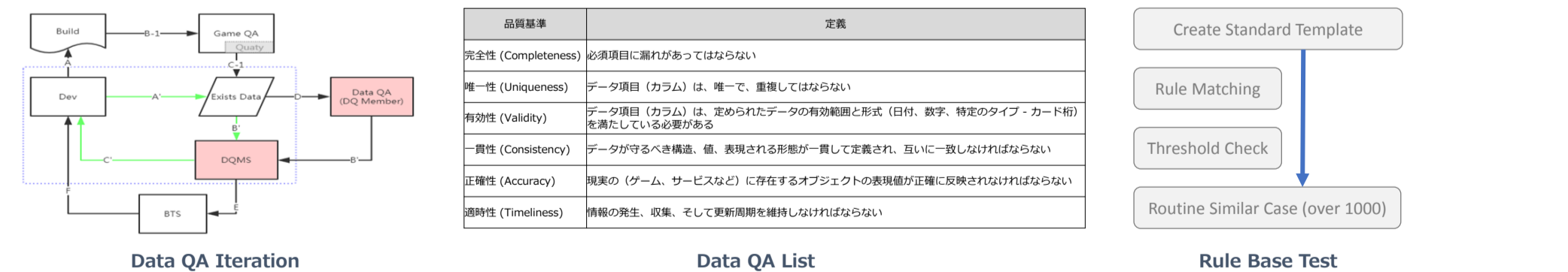
Sales Forecast Analysis

Dataのサービスにおいて、現在の状況が最も重要ですが、サービスの今後を予測し、方向性を確立することが必要となるため、コラボレーションを要求する場合は非常に多いです。そのような状況のために重要KPIに対する予測値を提供することになります。これを通じてより具体的な将来設計が可能になります。私たちは、これらのSales Forecastに対して様々な分析Open SourceやPackagesを考慮してきましたが、最終目的の方向を見つけるのは容易ではありませんでした。基本的なSeasonalityを確保するために月中の最初の日のみARIMAを導入することになりました。その理由としては、各サービスごとに月初KPI購入商品が月中に消費されるため、KPI同士の因果関係が維持されるからです。



Data Quality Management

Dataは保存のあと、多様に活用されるようになります。1次加工、2次加工などを経て、要約または分析結果を活用することになります。このような環境構築のために最も基本的なData Qualityが保証されていなければ、指標、分析、AIなどいかなる結果も保証することができなくなります。私たちは、これらのData Qualityについて様々なアプローチを通じ、Qualityを確保するための努力をしています。常に基本を保証しながら、未来のためのAI的DQMSを準備しています。



このようなコンテンツを通じてData-drivenを可能にすることができます。より詳しい内容についてご不明な点がある場合は、growthy_support@linecorp.comまでお問い合わせください。

We are also developing these various solutions

- > Analytics: User Behavior Dataに基づいた意思決定をサポートする分析サービス
- > Probe: ビックデータを照会/分析できるSelf分析サービス
- > ETL: データのETL (Extract, Transform, Load) を実行/管理するサービス
- > Catalog: GROWTHYのMeta情報を統合管理するサービス
- > Storage: ビックデータを収集、処理、保存するサービス