
influxdata
InfluxData のケーススタディ（サンプル版）
AN INFLUXDATA CASE STUDY
How Robinhood Built a Real-Time Anomaly Detection System to Monitor and Mitigate Risk
Robinhood がリスクを監視・軽減するために
リアルタイムで異常検知システムを構築した理由
External Contributors
Allison Wang
Software Engineer, Robinhood
Allison Wang（アリソン・ワン）
Robinhood ソフトウェアエンジニア
October 2019 (Revision 1)
2019 年 10 月

サンプル版の為、一部の記載を省略させていただきます。ご興味が合わる方は、日本代理店までご連絡をお願いいたします。

会社概要

ロビンフッドの物語はスタンフォード大学で始まりました。共同創設者の Baiju（バイジュ）氏と Vlad（ブラッド）氏は同大学のルームメイトで同級生でした。卒業後、彼らは荷物をまとめてニューヨークへ行き、2 つの金融会社を設立して、ヘッジファンドに自分たちのトレーディングソフトウェアを売り込みました。そこで彼らは、ウォール街の大企業が株取引の手数料を全くと言ってよいほど払っていないのに対して、ほとんどのアメリカ人は取引ごとに手数料を取られていることに気付きました。そこで彼らは、富裕層だけでなく誰もが金融市場を利用できるような金融商品を開発するために、カリフォルニアに戻ったのでした。

手数料無料投資のパイオニアであるロビンフッドは、あらゆる人のために金融を民主化することを使命としており、金融システムは誰にとっても機能するように構築されるべきだと信じています。ユーザーが自分のペースで投資を始められるような製品を同社が開発しているのは、このような背景があるからです。

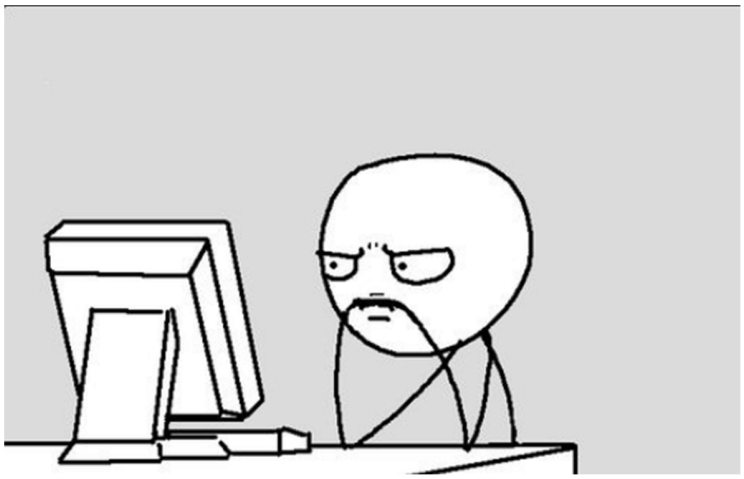
導入事例の概要

ロビンフッドは、ユーザーのスマートフォンやデスクトップ PC を利用した手数料無料の投資や取引機会を提供することで、金融システムの民主化を進めています。これはロビンフッドの社外から見ると非常にエキサイティングに思えますが、社内のチームではさまざまなリスク要因を把握して軽減するためのエンジニアリングソリューションを構築する必要がありました。リアルタイムのリスクモニタリングシステムを構築するために、ロビンフッドは InfluxDB（Go で書かれたオープンソースの時系列データベース）と Faust（Kafka ストリーム用のオープンソースの Python ストリーム処理ライブラリ）を選択しました。システムのベースとなるアーキテクチャには、時系列異常検知（InfluxDB）とリアルタイムストリーム処理（Faust/Kafka）の両方のセットアップが含まれています。

> 「時系列データが増加するにつれて、時系列の異常を把握・検知するのに必要な作業には多大のコストがかかるようになりました。そのため、何かがうまく機能していないときにインテリジェントに警告できる異常検知システムの構築を開始しました。」
>
> ソフトウェアエンジニア、*Allison Wang*（アリソン・ワング）氏

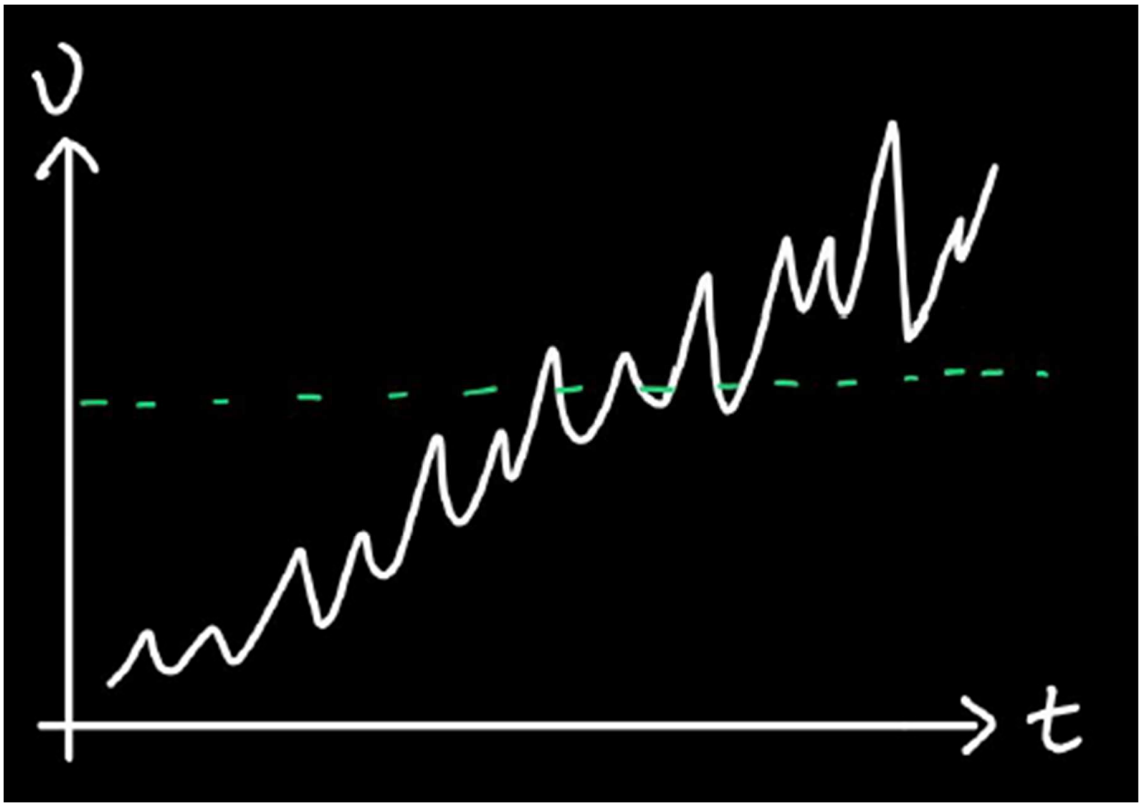
ビジネス上の問題点

リスクを監視して軽減するために、ロビンフッドは常に手動でダッシュボードを追跡する代わりに、重要なメトリクスに対してインテリジェントなリアルタイムアラートを設定したいと考えていました。大量のメトリクスに直面したエンジニアと IT 運用チームは、リスクを迅速に評価できるように、時系列の異常を検知するスマートで自動化された方法を求めていました。スクリーンを見つめて、何千もの時系列データを 24 時間 365 日追跡して、即座に行動を起こすことは現実的ではなく、拡張性もないので、ロビンフッドは異常検知システムを構築する必要に迫られていたのです。



技術的な問題

これらの問題を解決するために、ロビンフッドが最初に試みた異常検知ソリューションは、下図に示すように、基礎データがしきい値を超えるか下回るたびにアラートが発動されるしきい値ベースのアラートでした。



しきい値ベースのアラート

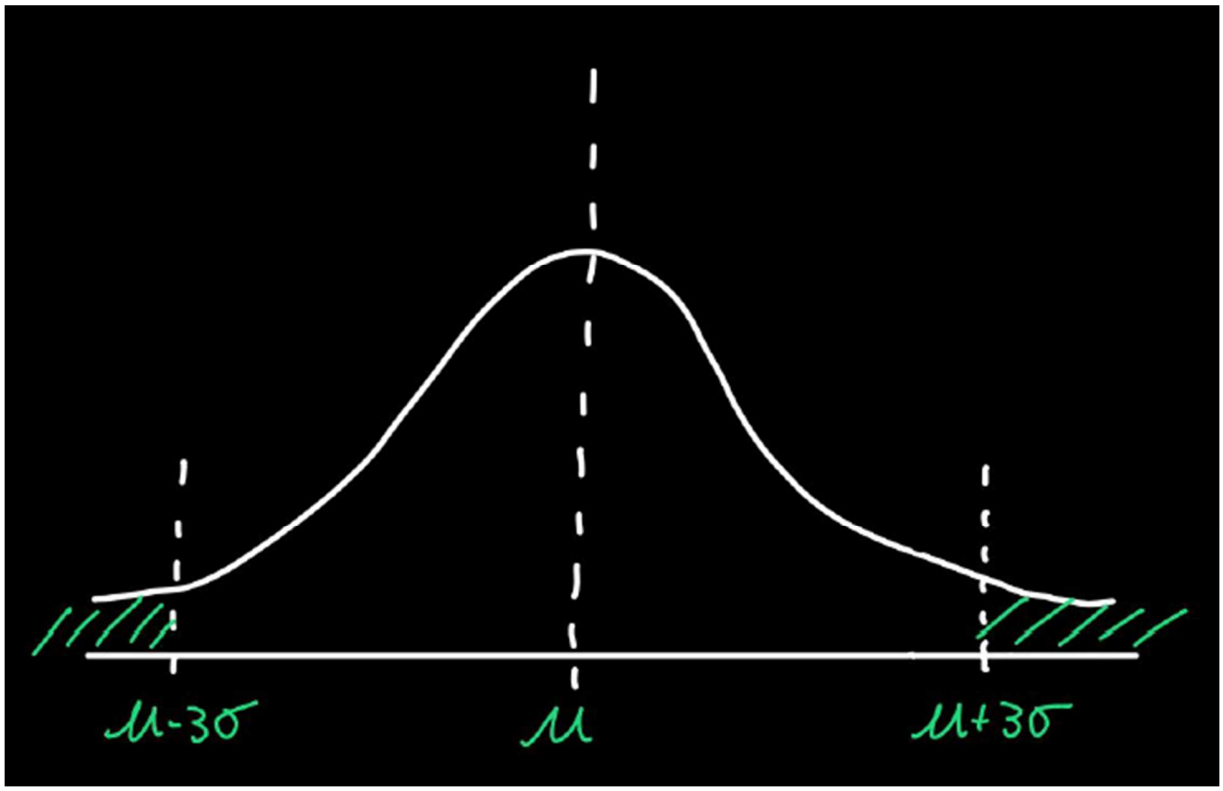
しきい値ベースのアラートは、単純な時系列であればうまく機能します（たとえば CPU 使用率が 80%を超えているか、または現在実行中のサーバーの割合が 100%を超えているかどうかを検知する場合）。しかし、季節性やトレンドを含んだ複雑な時系列に対応することはできません。上図に示すように、時系列データには上昇傾向があり、その中で何回も上下するパターンがあります。

固定しきい値を用いて異常をアラートする場合は、時系列データがしきい値を超えるとアラートをトリガーしますが、その後にしきい値より下がり、再びしきい値を超えるとうまく機能しません。したがって、このように複雑な時系列データの場合には、ダッシュボードを 24 時間 365 日チェックするのと同じ労力が要ることになります。このことからロビンフッドには異常検知アルゴリズムが必要になりました。

異常検知アルゴリズムの作成

同社は、異常状態を定義するために、履歴データを活用して着信データポイントを基に妥当なしきい値を決定しようとしました。そこで、正規分布と呼ばれる統計学に由来する概念に頼りました。正規分布の背後にある考え方は、データのリストがあればデータポイントの平均値と標準偏差を計算できるというものです。この正規分布の概念を活かして、ロビンフッドは 3 つの標準偏差外のデータに対してアラートを発することにしました。

- 1σ - 68.27%
- 2σ - 95.45%
- 3σ - 99.73%



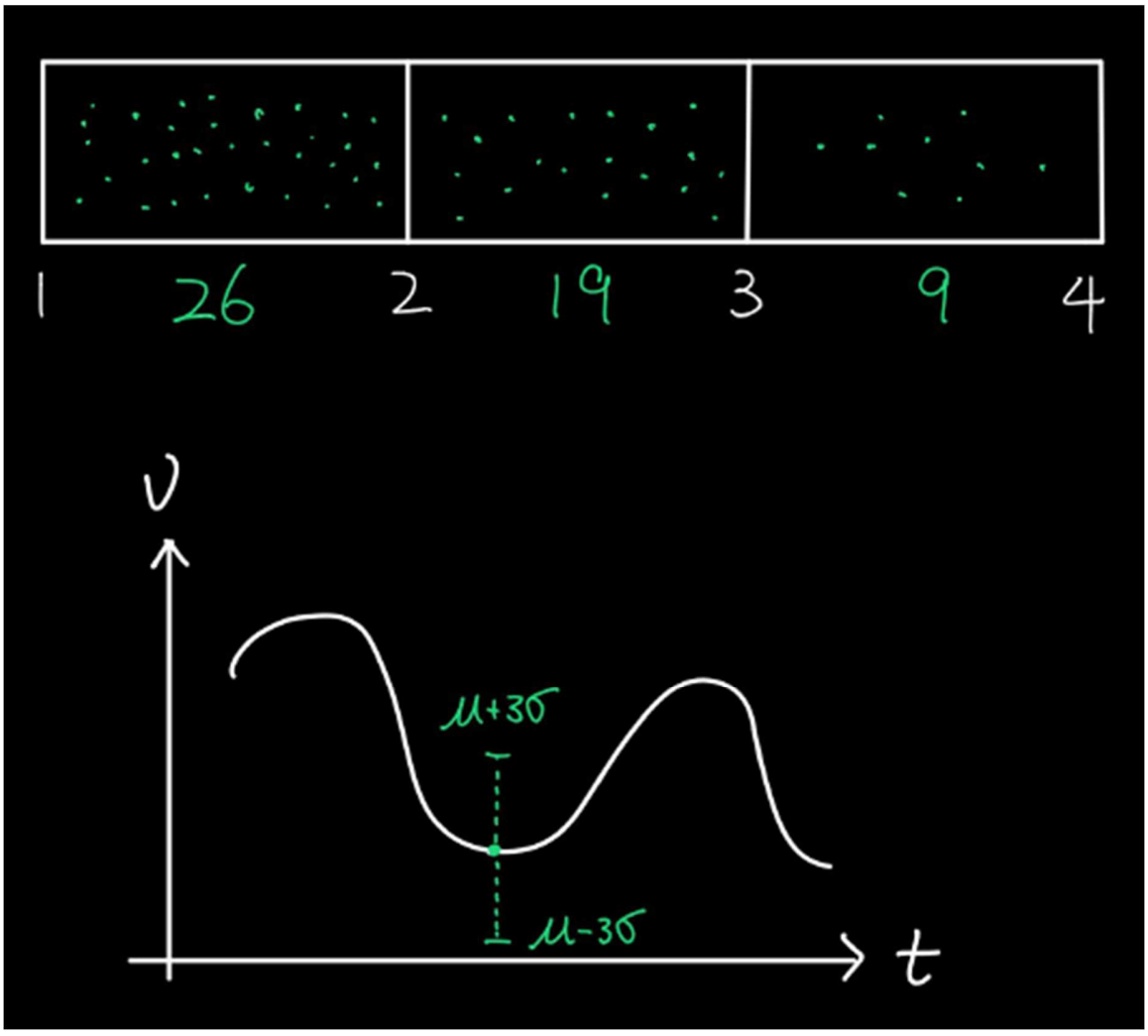
正規分布のデータ例。平均値から 3 標準偏差外にあるデータ（緑色の斜線）は、全データの 0.03%に過ぎない。

異常検知用のしきい値を標準偏差で定義すると（上図の例のような）非定常なデータの異常を検知しやすくなるため有利になります。標準偏差によって定義されたしきい値はデータのトレンドに従います。ロビンフッドは平均から 3 標準偏差外のものを異常と定義したので、データの 99.7%がこの範囲内に収まることになります。

データポイントのバケット化

正規分布の考え方を異常検知に応用する方法を次に示します。データストリームが入ってきたら以下を実行します。

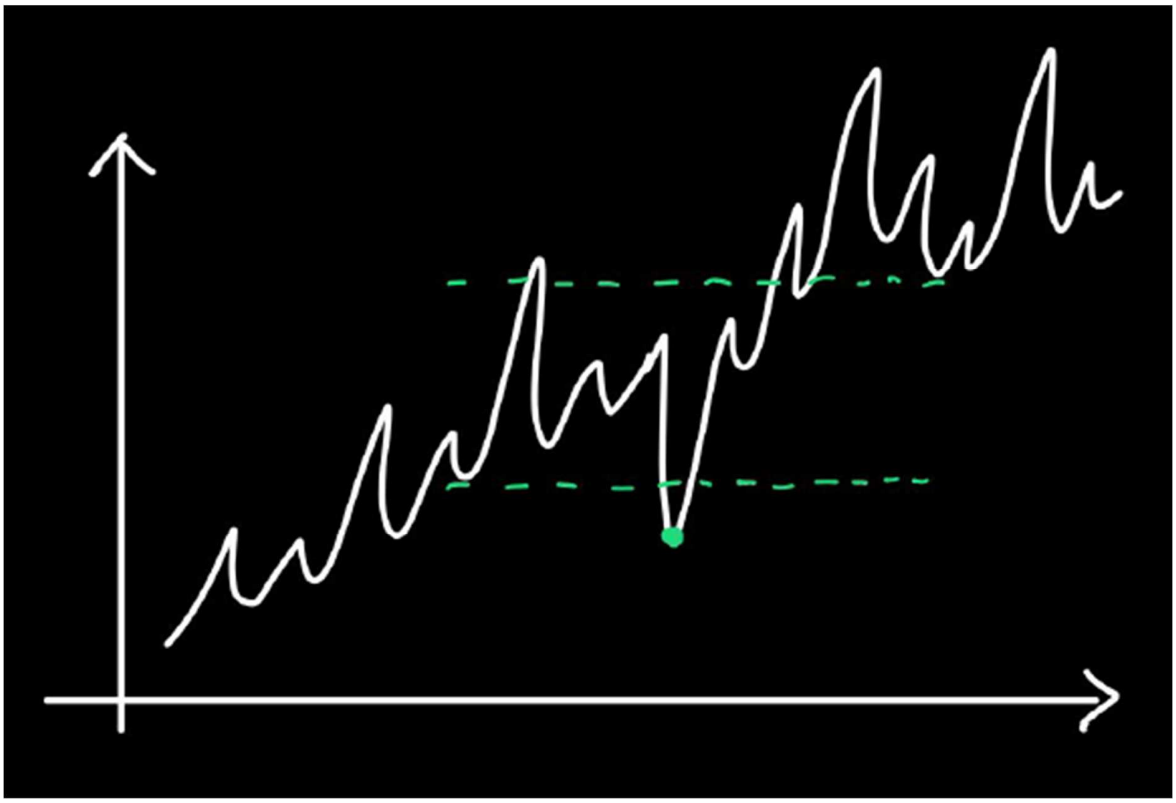
1. 1 分間など非常に短い間隔でデータポイントをバケット化する
2. 過去 30 日間にわたり 1 日の正規分布を毎分作成する



たとえば最初の 1 分と次の 1 分の間に着信するデータポイントの数（例：26）を数えます。次に 2 分目と 3 分目を数え、3 分目と 4 分目を数えるというように続けて行きます。集約が完了すると、過去 30 日間のデータポイントに適用します。こうして過去 30 日間にわたって 1 日に毎分集約されたデータポイントのリストができあがります。

集約した時系列データを利用して過去 30 日間の平均と標準偏差を計算し、それを時系列データのしきい値の境界線として使用します。しきい値を計算した後は、新しいデータポイントが着信するたびにその数値としきい値を比較することができます。もし着信したデータポイントがしきい値を超えているか下回っている場合はアラートが発出されます。これが過去データを使った異常検知の考え方です。

これで異常検知アルゴリズムを設定できたので、次のステップはアラートを自動化できるようにシステムでアルゴリズムを作成することでした。



着信データポイント（集約済）が（μ-3σ、μ+3σ）の範囲内にあるかどうかをチェックする

そこで時系列データを監視するには以下のシステム要件を満たす時系列データベースが必要となることに気付きました。

- 時系列データの取り込みと集約を高速に行うデータベース
- 異常をリアルタイムで照会し計算するシステム
- 可視化とアラーティングの機能

ソリューション

> 「*InfluxDB* には極めて優れたスタックが備わっており、システム構築時に必要なものをすべて提供してくれました。」

なぜ InfluxDB なのか？

サンプル版の為、省略

エンドツーエンドの異常検知システムの構築

サンプル版の為、省略

テクニカルアーキテクチャ

サンプル版の為、省略

データの取り込み

サンプル版の為、省略

可視化

サンプル版の為、省略

アラーティング

サンプル版の為、省略

ハイレベルシステムアーキテクチャ

サンプル版の為、省略

結果

> 「*InfluxDB* がなければこれほど短期間にシステムを構築し、本番環境でこれほどうまく稼働させることはできなかったでしょう。」

サンプル版の為、省略

InfluxData について

InfluxData は業界をリードする時系列プラットフォームである InfluxDB を開発した企業です。Cisco、IBM、Lego、Siemens、Tesla などの開発者や組織が、革新的な IoT、分析、監視アプリケーションを構築できるよう支援しています。当社の技術は、センサー、アプリケーション、コンピュータインフラストラクチャから生成される大量のタイムスタンプデータを処理するために構築されています。
InfluxDB は簡単に導入して拡張することができるため、開発者はアプリケーションの競争力を高める機能の開発に集中することができます。InfluxData はサンフランシスコに本社を置き、従業員は米国およびヨーロッパ全域で活躍しています。詳細については、www.influxdata.com にアクセスし、@InfluxDB をフォローしてください。

詳細はこちら

InfluxDB ドキュメント、ダウンロード、ガイド

InfluxDB をダウンロードする（英語）	https://www.influxdata.com/get-influxdb/
ドキュメントを入手する（英語）	https://docs.influxdata.com/
その他のケーススタディ（英語）	https://www.influxdata.com/_resources/
InfluxDB コミュニティに参加する（英語）	https://www.influxdata.com/community-showcase/

799 Market Street

San Francisco, CA 94103

(415) 295-1901

Web: https://www.InfluxData.com

Twitter: @InfluxDB https://twitter.com/influxdb?lang=en

Facebook: @InfluxDB https://www.facebook.com/influxdb/

日本代理店：

伊藤忠テクノソリューションズ株式会社

ＡＩビジネス部／ＤＳビジネス推進部

〒105-6950　東京都港区虎ノ門 4-1-1

神谷町トラストタワー

influxdb-sales@ctc-g.co.jp

InfluxDB 紹介ページ（日本語）：

https://www.ctc-g.co.jp/solutions/influxdb/

エンタープライズ版照会ページ（日本語）：

https://www.ctc-g.co.jp/solutions/influxdb/?menu=2#detail-top

事例資料ダウンロード・ページ（日本語）：

https://www.ctc-g.co.jp/solutions/influxdb/?menu=4#detail-top