rapidminer

プラットフォーム　セキュリティ

概要

WHITEPAPER

rapidminer.com

目次

企業のデータを最大限に活用する前に、データを扱う安全なプロセスを確立する必要があります。綿密に練られた機械学習プロジェクトであっても、ユーザーのアクセスを管理し経時的にデータを追跡するプロトコルがなければ、失敗してしまうかもしれません。

ここには、ツールの選択が大きく影響します。組織のデータを使用するすべてのテクノロジーは、セキュリティを確保するため、確立したプロトコルに準拠する必要があります。

以下の概要では、RapidMinerプラットフォームのセキュリティインフラをご紹介します。このインフラにより、コンプライアンスを気にする時間を減らし、データから実用的な知見を得る時間を増やすことができます。

概要

先進的な考えを持つ企業の多くは、自社のデータから膨大な価値を引き出せることを知っています。今日では、データは企業の現状や重要な指標に対するパフォーマンスを報告するだけではありません。データサイエンスを正しく実装出来れば、組織は顧客の行動に関する知見を得たり、リスク評価を行えたり、ビジネスの意思決定のありえる結果を予測することが可能になります。

データが新たな石油であると認識する企業が増えるにつれ、情報を収集し共有する速度が速くなっています。企業は、デジタルによる情報交換にますます依存するようになっています。知識の共有が速ければ速いほど、望ましいビジネスの成果をもたらす機会が多くなります。

「過去五年間で、主にIoTによって利用可能なデータ量が急激に増加しただけでなく、この生データの洪水を知見に変え、最終的にはアクションに変える新しいツールが開発されてきました。」

– ACHIEVING BUSINESS IMPACT WITH DATA,
マッキンゼー・アンド・カンパニー

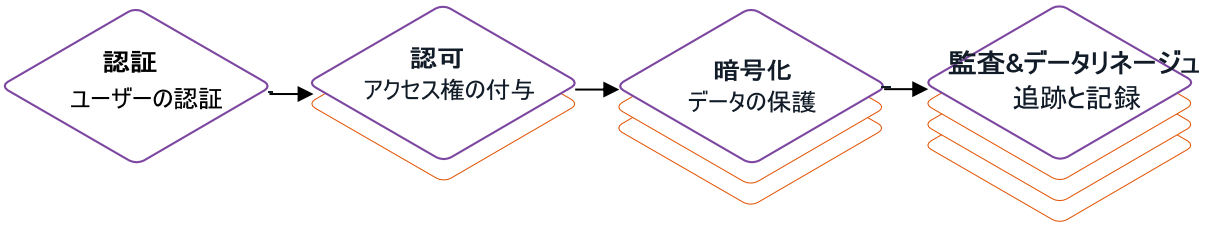
チャンスにはリスクが伴います。企業がデータから価値を引き出すことに注力する一方、悪意のある従業員や犯罪者はデータを盗み取り、悪用することに注力しています。近年、注目を集めているデータ侵害(データブリーチ)は顧客のデータを危険にさらし、多大な損害を被らせ、被害を受けた企業のブランドを傷つけています。

データサイエンスがもたらす競争上の優位を十分に活用する前に、大きなデータセットを扱うことに伴うリスクを理解し、積極的に対処していくことが重要です。以下のセクションでは、RapidMinerプラットフォームのセキュリティインフラを説明し、実用的な知見を得るためにセキュリティを犠牲にする必要はないことを示します。

企業のための多層セキュリティ

RapidMiner特有のセキュリティ対策について説明する前に、
RapidMinerでは4層のデータセキュリティ実装モデルに基づいて構築されていることを知っておいてください。

データセキュリティ実装モデル



データを扱うための安全なプロセスの開発には、二つの重要な領域に注目する必要があります。一つは、誰がデータにアクセスできるのかを完全にコントロールすること。もう一つは、データ自体を見えるようにすることです。両方の領域に取り組むことで、適切な人材に戦略的な決定を行わせることができ、これらの意思決定の背後にある情報を信用できるようになります。

RapidMinerプラットフォームのセキュリティ

RapidMinerのセキュリティプロトコルは、アカウント管理者やITスタッフに、ユーザーIDの認証、
権限の設定、企業データの暗号化、経時的な追跡を可能にします。
以下のセクションでは、この領域の概要と、これらの重要な領域の影響の大きさを説明し、これらと関連するRapidMinerのセキュリティ機能を紹介します。

- **ユーザーIDの認証**
- **ユーザーアクセスのコントロール**
- **企業データの保護**
- **データの追跡と記録**

ユーザーIDの認証

このセキュリティプロトコルはユーザー認証、すなわちユーザーが本人であることを確認する手段を意味しています。シングルサインオン(SSO)やSAMLなどの認証方法はインターネットに接続されている複数のデバイスからアプリにアクセスしたいという従業員の要望に応えるために、ますます一般的になってきています。

RapidMinerにおける認証

RapidMiner AI Hubのログインセキュリティを確保するために、プラットフォームの最初の認証レイヤーにKeycloakを取り入れました。これにより、認証と認可に標準プロトコルを使用しながら、RapidMinerプラットフォームの全コンポーネントでのシングルサインオン体験を可能にします。管理者はパスワードの複雑さの要件や二段階認証を設定することができます。

さらに、KeycloakはIDとユーザー連携にOpenID Connect、SAMLv2.0、LDAPなどの規格の使用を可能にし、全エンタープライズアプリケーション間でのSSO体験を提供します。

RapidMinerセキュリティフロー

RapidMinerにおける主な認証機能

- シングルサインオン/サインアウト(SSO)
- ソーシャルログイン (Google, GitHub etc.)
- 二段階認証 (2FA)
- SAML 認証サポート
- OAuth 2.0 サポート
- OpenID Connect (OIDC)

ユーザーアクセスのコントロール

データアクセスセキュリティにより管理者はネットワーク上でユーザーが見える(扱える)データの種類を定めることができます。このプロセスはセキュリティを向上させるだけでなく、従業員が自分の仕事に必要なものに集中できるようにし、必要でないものに気を取られずに済むようにします。

RapidMinerにおける認可

管理者はロールベースの権限を設定して管理者と他のユーザーの機能を分けることができ、またカスタムロールを作成してユーザーの作業できる範囲をプラットフォームの一部に限定することができます。

またKeycloakと統合することで、権限を作成し、認可ポリシーと結び付け、プラットフォーム内で認可の判断を実施することができます。ファイルレベルの権限スキーマを実装することで、管理者はRapidMinerのセントラルリポジトリに保存されているデータへのアクセスをコントロールできます。さらに、AI Hubではオープンで共同作業のしやすいプロジェクトで使用されるファイルへもアクセスを制御することができます。

最後に、ビッグデータについて、RapidMinerはHadoopクラスタで作成したどんなセキュリティポリシーも遵守し、ユーザーに定義した行や列レベルのフィルタリングまで下ることができます。

RapidMinerにおける主な認可機能

- OAuth 2.0 サポート
- OpenID Connect サポート
- Adminコンソールの組み込み
- トークンマッパー(Token Mappers)
- Hadoopセキュリティサポート
- グループ&ロールマッピング

本番環境で、RapidMiner管理者は各ユーザーにジョブの実行やスケジューリングを許可する特別な権限を割り当てることができます。

企業データの保護

セキュリティの観点では、暗号化は企業がデータを保護するための最も重要な手段の一つです。データの暗号化は、パスワードや復号キーを持つ人だけが閲覧できるようにデータを隠します。暗号化アルゴリズムは認証とデータの整合性を含むセキュリティ全体の様々な部分に影響を与えます。

RapidMinerにおける暗号化

RapidMiner AI HubはTLS(Transport Layer Security)プロトコルをサポートしており、ネットワーク上を移動するデータを暗号化します。接続メタデータも暗号化されるため、エンドユーザーやサードパーティに認証情報をさらすことなく、ビジネス上の重要なシステムへアクセスを可能にします。

またLDAP認証を有効にした組織は、LDAPプロパティを暗号化することで、設定の保護を強化できます。

最後に、アプリケーションレベルではリポジトリ全体は暗号化されていないことに注意してください。
RapidMinerはファイルシステムレベルでの暗号化です。

RapidMinerにおける主なデータ保護機能

- TLS (Transport Layer Security) サポート
- LDAPプロパティの暗号化
- Hadoop HDFS 暗号化

データの追跡と記録

管理者はデータアクセスがどこに与えられていても(例えばデータベース、サーバー、Hadoop クラスタなど)、監視し、監査を行うことができます。これは、セキュアなデータサイエンスを行うための重要な要素です。データリネージュを理解することは非常に重要です。意思決定に使用されるデータを信頼するには、組織は最初にデータの起源を理解し、時間の経過とともに変化するデータを把握できるようにならなければなりません。

管理者コンソール

オートロギング

ロールバック機能

RapidMinerにおける監査とデータリネージュ

管理者が経時的にプロセスとモデルを監査できるように、RapidMinerはロギングとバージョン管理機能を提供しています。ログは全製品で生成され、管理者はどのユーザーが、いつログインしたのかを確認できます。RapidMiner AI Hubでの認証と認可に関する全ての変更は、監査ログで追跡されます。

またAI Hubはデータリネージュを追跡できるように設計されています。プロセスにおいて、メタデータはモデルがどこで作成されたのかだけでなく、データの一部がどのように変換されたのかも示します。またプロジェクトにはコミットの履歴があり、誰がいつデータに変更を加えたのかを確認できます。

RapidMinerにおける主な追跡と記録機能

- バージョン管理とロールバック機能
- オートロギング
- 管理者コンソールの組み込み
- イベントマッピング
- 管理者のセッション管理機能
- グループ&ロールマッピング

まとめ

組織が強力なデータサイエンスを実行して利益を得るには、データを扱うセキュアなプロセスを構築することが非常に重要です。最も強いプロセスは、ユーザーの認証、アクセスレベルのコントロール、データの保護、データリネージュの理解から成り立っています。

RapidMinerは企業にこれらのことを可能にします。誰がプラットフォームにアクセスできるのか、何を見れるのかを組織がすべてコントロールできることで、RapidMinerのセキュリティプロトコルは許可された関係者にしか機密情報が見えないことを保証します。

また情報の完全性を維持するために、ユーザーが時間の経過とともに行われる変更を追跡できる機能もあります。

セキュアなプロセスが整えば、組織内の従業員は実用的なビジネスの知見の収集や、その知見を活かすことに集中することができます。

RapidMinerは誰もが未来を積極的に形作る力を持てるように、エンタープライズAIを改革し続けています。企業内のどんなスキルレベルの人でも、すばやくAIソリューションを作成し、運用できるようにすることで、ビジネスインパクトをすぐにもたらすことができます。データ前処理、機械学習、モデル運用まで全てを備えたプラットフォームを提供し、データサイエンティストには深さを、それ以外の人には複雑なタスクを簡単にするユーザー体験を提供します。RapidMinerのCoE(Center of Excellence)とRapidMiner Academy は顧客の経験やリソースのレベルにかかわらず、成功を保証します。150か国を超える40,000以上の組織が収益の増加、コストの削減、リスクの軽減にRapidMinerを利用しています。

主要機能のまとめ

認証

- シングルサインオン/サインアウト(SSO)
- ソーシャルログイン (Google, GitHub etc.)
- 二段階認証 (2FA)
- SAML 認証サポート
- OAuth 2.0 サポート
- OpenID Connect (OIDC)

認可

- OAuth 2.0 サポート
- OpenID Connect サポート
- Adminコンソールの組み込み
- トークンマッパー(Token Mappers)
- Hadoopセキュリティサポート グループ&
- ロールマッピング

データの保護

- TLS (Transport Layer Security) サポート
- LDAPプロパティの暗号化
- Hadoop HDFS 暗号化

監査とデータリネージュ

- バージョン管理とロールバック機能
- オートロギング
- 管理コンソールの組み込み
- イベントマッピング
- 管理者のセッション管理機能
- グループ&ロールマッピング