

Synthèse de l'article : *Application des technologies Big Data dans le processus aval de la sidérurgie*

I. Chibane et S. Fellah
Auteurs originaux : F. Avelino et Al

Résumé : L'évolution des techniques et des infrastructures de collecte de données dans le laminage d'acier a permis d'accumuler des volumes de données considérables. Rendant leurs exploitations pour différents cas d'utilisation un véritable challenge. C'est pourquoi en 2018 fut lancé le projet RFCS NewTech4Steel utilisant des concepts de Big Data dans l'optique d'utiliser ces quantités massives de données pour améliorer la stabilité du processus et la qualité du produit dans la production d'acier. Étudiant le cas de l'usine Marcegaglia Ravenna, la solution s'inspirant de l'architecture pour le traitement des données, et l'utilisation d'algorithme de machine learning afin d'optimiser la productivité globale. Le projet répond aux besoins de l'usine en fournissant des prédictions rapides des défauts et des paramètres de processus en temps réel s'intégrant parfaitement dans les systèmes informatiques existants.

1 Introduction

Le domaine du laminage ayant une grande influence dans la qualité finale des produits dans la production d'acier, il est important de définir les relations entre les caractéristiques de qualité et les conditions dynamiques du processus. Malheureusement, il existe un décalage entre les équipements de haute technologie avec des données à haut débit et les méthodologies d'exploitation actuellement utilisés. C'est pourquoi le projet NewTech4Steel introduit des technologies de Big Data et de machine learning pour extraire automatiquement et efficacement des relations. Le projet vise à améliorer la prédiction des aspects cruciaux de la stabilité du processus et de la qualité du produit. Les objectifs technologiques incluent l'amélioration des solutions de stockage, l'intégration d'outils innovants de la technologie Big Data, et l'application des dernières technologies d'analyse de données. L'accent est mis sur l'adaptation de ces technologies aux environnements industriels existants, la connexion aux systèmes existants et les exigences en matière d'interfaces homme-machine. Le projet évalue les systèmes à travers des indicateurs de performance et estime les impacts potentiels sur les cas industriels étudiés. On peut noter l'existence de plateforme de traitement de données dans d'autres industries, mais qui souffre d'un manque de généralisation les rendant difficiles à adapter, mais dont ce projet s'inspire quand même sur certains points.

2 Évaluation Industrielle

L'objectif du projet est de résoudre les problèmes de l'industrie de la production d'acier en utilisant le big data et l'apprentissage automatique. Il consiste à extraire, traiter et évaluer des ensembles de données de manière rapide et efficace pour alimenter des modèles de prédiction et de benchmarking. La partie technique implique la collecte de données par des capteurs, leur stockage via des technologies de big data telles que les bases de données NoSQL, la communication fluide entre les systèmes et modèles grâce à l'architecture lambda pour le prétraitement en temps réel, tout en utilisant simultanément les données historiques du système. Cela permettant d'optimiser les algorithmes d'analyse et de prédiction, fournissant des résultats communiqués aux systèmes et visualisables via une IHM. Les applications incluent le contrôle qualité en temps réel, la maintenance prédictive, et l'optimisation des chaînes d'approvisionnement ainsi que des paramètres du processus.

3 Architecture de la Plateforme

L'architecture logicielle mise en place pour notre cas d'utilisation se constitue d'Apache Nifi pour l'ingestion de données, Apache Hadoop pour le stockage, Apache Spark pour le traitement distribué, et Keras pour l'apprentissage automatique. Les données étant collectées à partir de systèmes de traitement industriels variés, stockées dans un datalake, et visualisées via une application web dédiée.

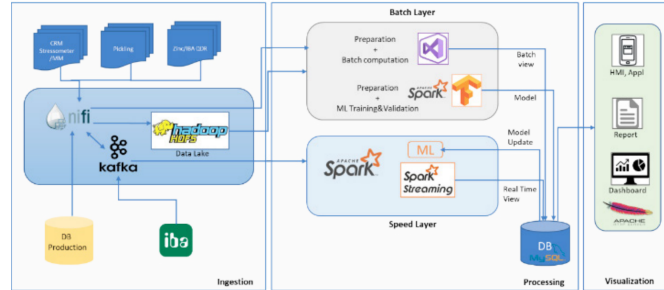


Figure 1: Architecture de la plateforme

3.1 Acquisition des Données

L'étape préliminaire du projet, et l'une des plus importantes, consiste en l'acquisition et la collection des données de manière robuste et fiable. Pour cela, différents équipements et capteurs situés dans les unités de l'usine de production sont utilisés permettant de mesurer et capturer les informations en temps réel. Les données collectées sont majoritairement des mesures numériques liées au processus de production permettant par la suite le contrôle et l'optimisation de la production, l'entraînement de modèles prédictif via le machine learning et la maintenance préventive. Les données sont ensuite stockées et distribuées de manière continue en utilisant des outils logiciels tels que Kafka, mais aussi Apache NIFI pour gérer le flux de données entre les différentes étapes du processus.

3.2 Stockage des Données

Pour le stockage des données, nous utilisons comme type de stockage principal un datalake sous Apache Hadoop HDFS, permettant de stocker différent type de données brutes ou des données traitées, et cela, sans contrainte. Nous utiliserons aussi comme stockage secondaire la base de données relationnelle MySQL étant très performante pour gérer des nombres importants de transaction, cela permettant de stocker des tables de configuration statique nécessaires au fonctionnement du post-traitement et de la visualisation, mais aussi les différents batches de données reçus en continu.

3.3 Analyse et Prétraitement des Données

Pour analyser les dépendances entre les paramètres de processus et les objectifs d'utilisation, nous combinons des outils statistiques et d'apprentissage automatique dans Spark (Python) et C++, en commençant par le prétraitement des données pour assurer leur fiabilité. Nous créons et sélectionnons des caractéristiques, réduisons la dimensionnalité via l'ACP, et utilisons des réseaux de neurones artificiels (ANN) dans Keras pour la prédiction. Cette méthode inclut des schémas d'ANN variés pour les prédictions basées sur les bobines et les positions spécifiques, avec un accent sur la précision et l'efficacité en temps réel grâce à un système de streaming pour les données en ligne.

3.4 Pipeline et Workflow de la Plateforme

Dans le cadre du projet, deux pipelines Big Data principaux sont déployés. Le premier est destiné au traitement par lots pour la classification de la criticité des bobines dans le processus du laminage à froid. Ce pipeline, orchestré sous Apache NIFI, ingère des données de diverses sources, les stocke sur HDFS, puis les combine et les intègre pour l'analyse. Un modèle d'apprentissage automatique basé sur des réseaux de neurones artificiels (ANN) est ensuite utilisé pour prédire les conditions des

paramètres cibles le long de la bobine. Les résultats, y compris l'indice global de criticité de la bobine, sont stockés dans MySQL pour la visualisation.

Le second pipeline concerne la prédiction en temps réel pendant la phase de galvanisation des bobines. Il traite les données des paramètres de processus, capturées et transmises par ibaPDA-Data-Store-Kafka et stockées sous forme de messages JSON sur Kafka. Ces données sont ensuite traitées par une couche de vitesse utilisant PySpark Structured Streaming, qui exécute le modèle ANN pour prédire les conditions de traitement spécifiques. Les résultats sont directement enregistrés dans la base de données MySQL, permettant une visualisation en ligne dans l'interface homme-machine (IHM).

3.5 Visualisation des données

La visualisation est un aspect clé du projet NewTech4Steel. Pour cela, une application web spécifique a été développée, permettant aux opérateurs de l'usine et au département Qualité et Contrôle des Processus d'accéder aux informations importantes, notamment pour la classification des bobines en lots. Cette application, basée sur Apache HTTP Server et PHP, récupère les données de la base de données MySQL pour offrir une vue d'ensemble des résultats des lots et la criticité des bobines sur la ligne de galvanisation (un processus de revêtement métallique pour prévenir la corrosion). La criticité est visualisée par la couleur dans l'interface, indiquant les risques potentiels de la prochaine étape de galvanisation. En cliquant sur une bobine spécifique, l'utilisateur peut voir les détails des configurations prédites pour chaque section critique de la bobine. En outre, l'application offre une visualisation en temps réel du traitement spécifique d'une bobine en cours de galvanisation, avec des mises à jour continues des valeurs de criticité liées aux dispositifs de centrage et de guidage.

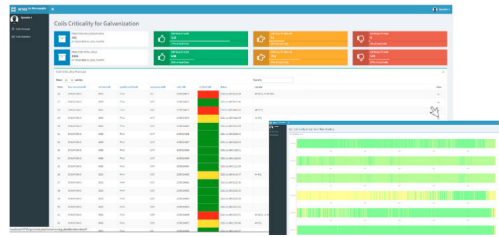


Figure 2: IHM de la plateforme

4 Déploiement

Parmi les différentes méthodes de déploiement existantes (Externalisée, sur site, cloud), un déploiement en interne donc sur site fut choisi en raison des besoins industriels. Un serveur connecté à l'infrastructure informatique de l'usine fut utilisé et où les différents outils utilisés tels que les logiciels Apache, le framework .Net, les bibliothèques Python pour le Machine Learning, et le SGBD MySQL y ont été installés. Tirant profit de leurs natures Open Source pour faciliter leurs implémentations. Un prototype aura été testé pour différents types d'opération sur les données (collecte, transfert, traitement et visualisation) ainsi qu'évaluer pour voir sa capacité à gérer l'ingestion de lots de données et le streaming de données en temps réel étant des cas courants rencontrés en industrie.

5 Conclusion

Cet article présente une infrastructure avancée intégrant des outils de Big Data et des modèles IA/ML, conçue pour optimiser le processus de production sidérurgique. Le système, s'appuyant sur des modèles de réseaux de neurones artificiels (ANN), permet une détection préventive de la criticité des produits et assure une prédiction en temps réel des conditions de processus à risque. Les résultats obtenus apportent des informations clés pour l'amélioration continue de la production. Le projet a permis au partenaire industriel d'exploiter plus efficacement les données disponibles, améliorant ainsi significativement la qualité du processus de production et du produit final.