# Machine Learning Model Deployment

Abstract

This project leverages IBM Cloud Watson Studio to develop and deploy predictive models for customer churn analysis in the telecommunications industry. By collecting and analyzing customer data, training machine learning models, and integrating them into operational systems, the project aims to proactively reduce churn rates and enhance customer retention strategies.

Problem statement

Become a wizard of predictive analytics with IBM Cloud Watson Studio. Train machine learning models to predict outcomes in real-time. Deploy the models as web services and integrate them into your applications. Unlock the magic of data-driven insights and make informed decisions like never before!

Problem definition

The project involves training a machine learning model using IBM Cloud Watson Studio and deploying it as a web service. The goal is to become proficient in predictive analytics by creating a model that can predict outcomes in real-time. The project encompasses defining the predictive use case, selecting a suitable dataset, training a machine learning model,deploying the model as a web service, and integrating it into applications.

Steps of ML Model Creation
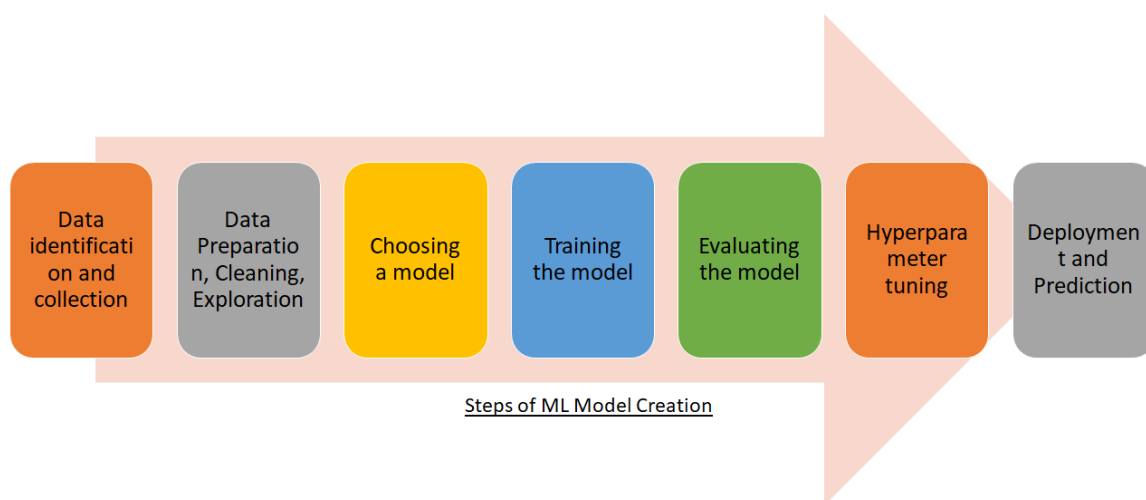
Design thinking

PredictiveUse Case: Automated loan approval system

Objective:

To build an automated loan approval system using machine learning that assesses loan applicants' creditworthiness based on their credit history and other relevant factors, allowing a financial institution to make more accurate and efficient lending decisions.

Dataset Collection

1. Identify Data Sources:

You'll need to identify data sources that contain relevant information for assessing credit risk. Potential sources include:

Credit bureaus: They provide credit reports and scores for individuals and businesses.

Financial institutions: Loan application data, payment history, and financial statements.

Employment records: Information about an applicant's job history and income.

Public records: Data on bankruptcies, tax liens, and other financial events.

Customer surveys: Additional data, such as personal details and preferences.

2. Data Privacy and Compliance:

Ensure that you comply with data privacy regulations (e.g., GDPR, HIPAA) when collecting and handling sensitive customer information.

Anonymize or pseudonymize personally identifiable information (PII) to protect customer privacy.

3. Data Elements:

Collect a wide range of data elements that can be relevant for assessing credit risk. These may include:

Applicant demographics: Age, gender, address, etc.

Credit history: Credit scores, credit utilization, and payment history.

Financial information: Income, debt levels, and assets.

Loan details: Loan amount, interest rate, loan purpose, and term.

Employment history: Job stability and income source.

Public records: Bankruptcies, tax liens, and judgments.

4. Data Quality Assurance:

Perform data quality checks to address missing values, outliers, and data inconsistencies.

Validate data against known standards and perform data cleansing if necessary.

5. Feature Engineering:

Create relevant features from the collected data. For instance, calculate debt-to-income ratios or derive new features that may be indicative of creditworthiness.

6. Data Splitting:

Divide the dataset into training, validation, and test sets. The training set is used for model training, the validation set for model tuning, and the test set for final model evaluation.

7. Data Documentation:

Create comprehensive documentation that describes the dataset's schema, data sources, preprocessing steps, and any transformations applied. This documentation will be valuable for future reference.

8. Data Storage and Security:

Ensure that the collected dataset is securely stored, and access is restricted to authorized personnel only.

9. Update and Maintenance:

Consider how you will update and maintain the dataset over time as new data becomes available. Regular updates can improve the model's accuracy and relevance.

Model Training

For an automated loan approval system, you need a model training algorithm that can effectively classify loan applications into approved or rejected categories based on various applicant attributes and credit-related features.Here we can use two algorithms namely Logistic Regression,Random Forest or Gradient Boosting.

When choosing between these algorithms for your automated loan approval system, consider the following factors:

Regulatory Requirements: If your system needs to meet strict regulatory compliance and transparency standards, logistic regression may be a preferable choice due to its high interpretability.

Model Complexity: If your dataset contains intricate interactions and non-linearities, ensemble methods like Random Forest or Gradient Boosting may provide better predictive performance.

Data Size: For small to moderately sized datasets, logistic regression is often sufficient. For larger datasets with complex patterns, ensemble methods can be more effective.

Speed and Efficiency: Logistic regression typically trains faster and requires fewer computational resources compared to ensemble methods, which can be important for real-time or high-throughput loan approval systems.

Model Explainability: Consider how important it is to provide explanations for loan approval decisions. Logistic regression offers straightforward explanations, while ensemble methods are more complex to interpret.

Model Deployment:

Select a Deployment Platform: Choose a suitable platform for deploying your machine learning model. Common options include cloud-based services like AWS, Azure, or Google Cloud, or deployment platforms offered by machine learning frameworks like TensorFlow Serving or scikit-learn.

Package Your Model: Prepare your trained model for deployment. This typically involves saving it in a format that the chosen deployment platform supports (e.g., TensorFlow SavedModel, scikit-learn joblib, or ONNX for interoperability).

Set Up Deployment Environment: Create a deployment environment that includes the necessary dependencies and libraries to run your model. Docker containers are often used to encapsulate the environment and ensure consistency.

Expose an API: Create an API (Application Programming Interface) to serve your model. This API will accept loan application data as input and return loan approval predictions as output. Common API protocols include RESTful APIs or gRPC.

Integration with Loan Approval System:

Data Ingestion: Ensure that your loan approval system can efficiently ingest loan application data. This data should be in a format compatible with your deployed model's input requirements.

API Integration: Integrate the API endpoint (created during model deployment) into your loan approval system's architecture. This may involve making HTTP requests to the API endpoint to get predictions.

Feedback Loop: Establish a feedback loop to continuously improve the model's performance. Collect information about approved and rejected loans, model predictions, and actual outcomes. Use this data for model monitoring and retraining.

Advantages of this project

Efficiency and Speed:

Automated loan approval significantly reduces the time required to process loan applications. This results in faster decisions and improved customer satisfaction.

Consistency:

Machine learning models make lending decisions based on predefined criteria, ensuring consistency and reducing the potential for human bias or error.

<u>Improved Accuracy</u>:

Machine learning models can analyze a wide range of applicant attributes and credit features, leading to more accurate credit risk assessments and reduced default rates.

<u>Enhanced Risk Management</u>:

By using data-driven models, financial institutions can better manage and mitigate credit risk, resulting in reduced financial losses.
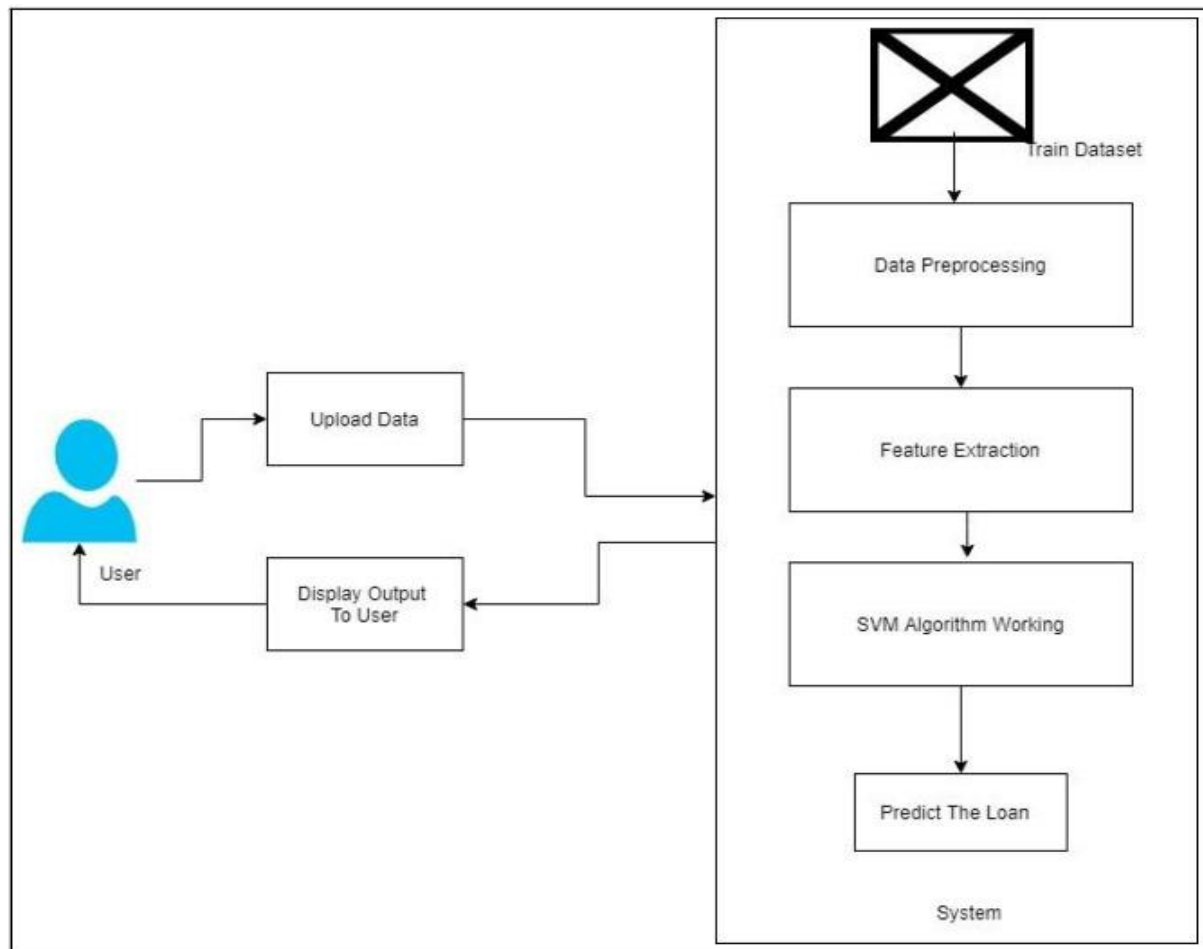
Conclusion:

In conclusion, the implementation of an automated loan approval system utilizing machine learning represents a transformative leap forward in the field of lending and credit risk assessment. This project demonstrates the fusion of cutting-edge technology with financial services, providing numerous benefits to financial institutions, customers, and the broader lending ecosystem.