Linear Regression in Azure ML: A Practical Example

This presentation will guide you through a real-world application of linear regression using Azure ML to predict credit limits based on user demographics and credit history.

M **by MBA IQAC**

Dataset Overview

Features

The dataset includes a range of features, both categorical and numerical. These include age, working status, credit history, and more. The features are carefully selected to influence credit limit.

Target Variable

The target variable is credit limit, representing the maximum amount of credit a user is eligible for. This is the value we aim to predict using linear regression.

Data Preprocessing: Transforming Data

Metadata

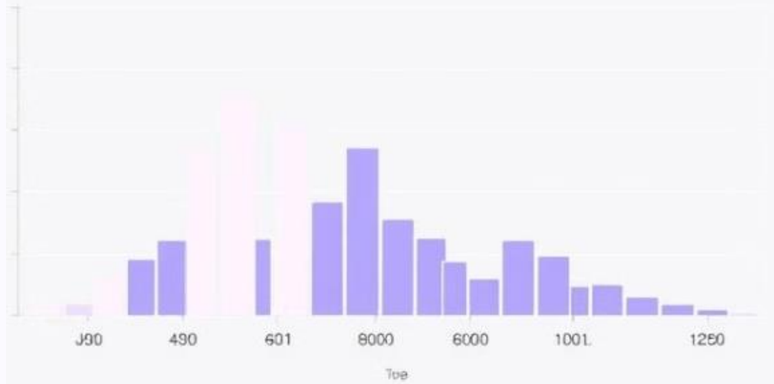
First, we identify and categorize features as numerical or categorical. This helps select appropriate transformations for each.

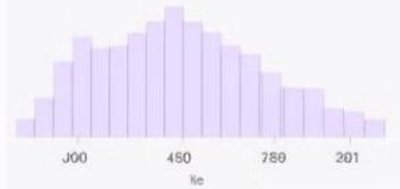
Log Transformation

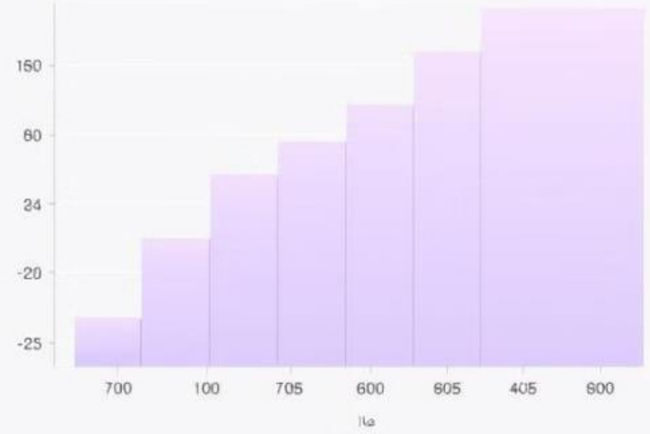
To normalize large credit limit values, we apply a log transformation. This ensures that the model is not overly influenced by outliers.

Normalization

Next, we apply z-score normalization. This standardizes the data, ensuring that features with different units or scales do not bias the model.







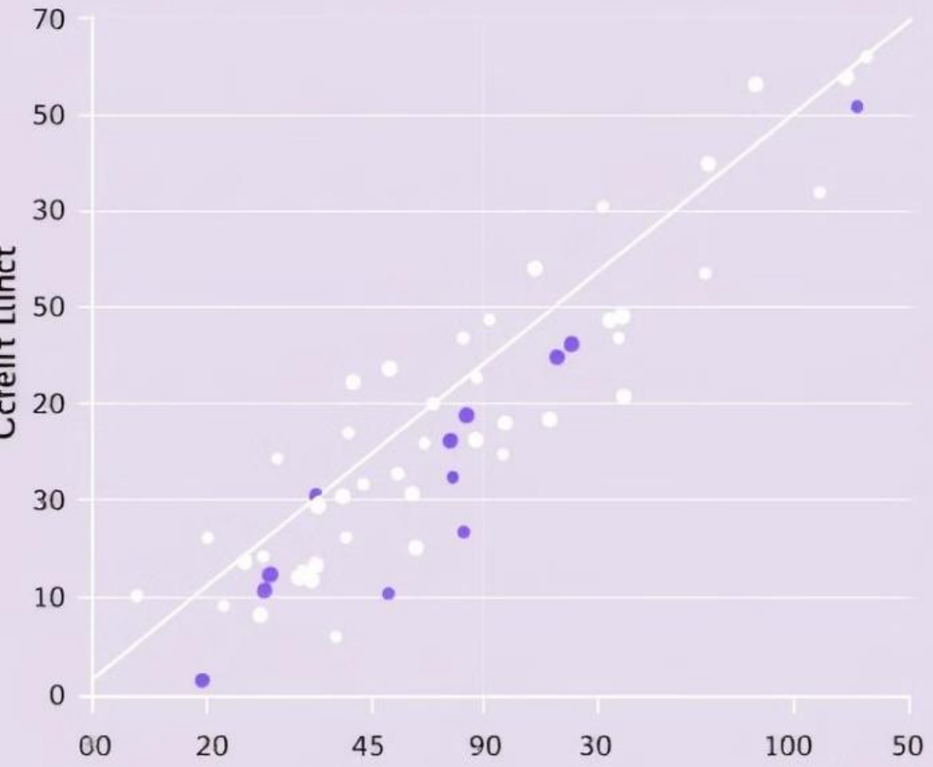
Model Training

Training Set

The data is split into a 70% training set and a 30% test set. The training set is used to train the linear regression model.

Linear Regression

We use linear regression to model the relationship between the transformed features and the credit limit. The model learns the coefficients for each feature.



Model Evaluation

1. After making predictions on the test set, we reverse the log transformation on the credit limits. This allows for a direct comparison with actual values.

2. Root Mean Squared Error (RMSE) is used to measure the model's accuracy. It represents the average difference between the predicted and actual credit limits.

Visualizing the Model

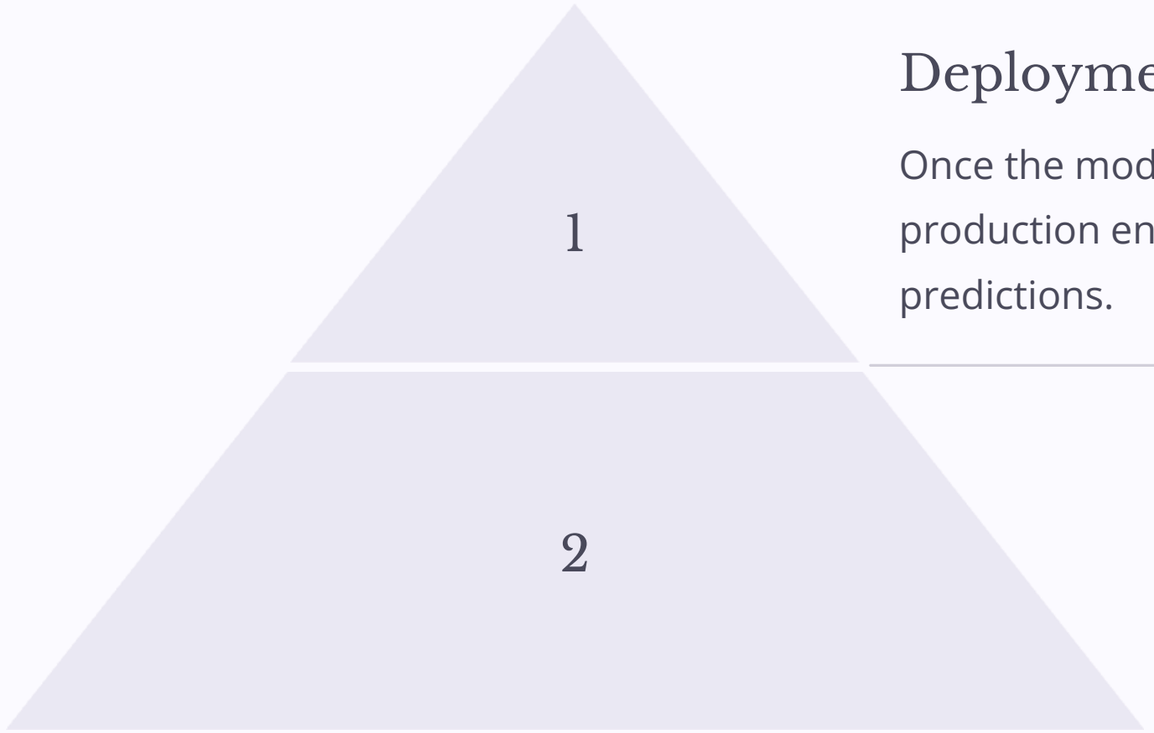
Data Exploration

Power BI dashboards are used to visualize the actual vs. predicted credit limits, providing insights into the model's performance and accuracy.

Model Interpretation

By analyzing these visualizations, we gain a deeper understanding of the model's predictions and how well it captures the underlying patterns in the data.

Model Deployment and Use



Deployment

Once the model is evaluated and deemed satisfactory, it is deployed to a production environment. This makes the model available for real-time predictions.

Predictive Capabilities

New user data can then be fed into the deployed model, allowing for accurate predictions of credit limits based on the trained model.

Key Takeaways and Next Steps

1

Linear Regression Power

Azure ML provides a powerful platform for building and deploying linear regression models to predict credit limits effectively.

2

Data Preprocessing Importance

Data preprocessing plays a crucial role in model accuracy. It ensures that the model receives clean and properly transformed data.

3

Continuous Improvement

Model evaluation and visualization are essential for continuous improvement. Regular monitoring and updates ensure model effectiveness.