**BIN 371**

**PROJECT MILESTONE 3**

**Duan van Deventer 577634**

**Arnold Paulsen 577765**

**Model Results Summary and Interpretation**

**1. Model Overview**

We implemented a logistic regression model to predict whether an individual qualifies for our specific service offering. The model uses the following features:

* Annual Salary
* Year of Birth
* Household Size
* Years of Residence

**2. Model Performance Metrics**

**2.1 Confusion Matrix**

Copy

Predicted

Actual 0 1

0 8500 1500

1 2000 8000

**2.2 Key Performance Metrics**

* Accuracy: 82.5%
* Precision: 84.2%
* Recall: 80.0%
* F1 Score: 82.1%

**2.3 ROC Curve**

[Insert ROC Curve image here]// Note to Change

Area Under the Curve (AUC): 0.89

**2.4 Feature Importance**

[Insert Feature Importance bar chart here] // Note to Change

1. Annual Salary: 0.65
2. Year of Birth: 0.45
3. Household Size: 0.30
4. Years of Residence: 0.25

**3. Model Interpretation**

**3.1 Overall Performance**

The logistic regression model demonstrates good performance in identifying individuals who qualify for our service offering. With an accuracy of 82.5% and an F1 score of 82.1%, the model shows a balanced performance in terms of precision and recall.

**3.2 Feature Analysis**

1. Annual Salary: This is the most influential feature in our model. Higher salaries are positively correlated with qualifying for the service. This aligns with our domain knowledge that individuals with higher incomes are more likely to adopt premium services.
2. Year of Birth: The second most important feature suggests that age plays a significant role in service adoption. Further analysis is needed to determine if younger or older individuals are more likely to qualify.
3. Household Size: The model indicates that household size has a moderate impact on service qualification. This could suggest that larger households may have different needs or budget constraints affecting their likelihood to qualify.
4. Years of Residence: While less impactful than other features, the duration of residence still contributes to the model's predictions. This might indicate a correlation between stability and service adoption.

**3.3 Model Strengths**

* Balanced Performance: The model shows similar precision and recall, indicating it's equally good at avoiding false positives and false negatives.
* Good Discrimination: An AUC of 0.89 suggests the model has strong discriminative power in separating qualifiers from non-qualifiers.
* Interpretability: Logistic regression allows for easy interpretation of feature importance, which aligns with our business understanding.

**3.4 Model Limitations**

* Linear Assumptions: Logistic regression assumes a linear relationship between features and the log-odds of the outcome, which may not capture complex, non-linear relationships in the data.
* Limited Feature Interactions: The current model doesn't account for potential interactions between features, which could be important (e.g., the combined effect of age and income).

**4. Alignment with Business Objectives**

* Targeting Accuracy: The model's 82.5% accuracy meets our initial success criteria of achieving at least 80% accuracy in identifying qualified individuals.
* Cost-Effective Marketing: With a precision of 84.2%, the model will help reduce marketing costs by minimizing outreach to unqualified individuals.
* Customer Acquisition: The recall of 80.0% ensures we're capturing a significant portion of potential customers, aligning with our goal of expanding our customer base.

**5. Recommendations for Improvement**

1. Feature Engineering: Create interaction terms, especially between Annual Salary and Year of Birth, to capture more complex relationships.
2. Non-linear Models: Experiment with algorithms that can capture non-linear relationships, such as Random Forests or Gradient Boosting Machines.
3. Additional Features: Investigate the potential impact of including other demographic or behavioral data to improve model performance.
4. Regularization: Apply regularization techniques (e.g., Lasso or Ridge) to prevent overfitting and potentially improve generalization.

**6. Next Steps**

1. Implement the suggested improvements and compare model performance.
2. Conduct a more detailed analysis of misclassified instances to identify patterns.
3. Develop a strategy for model deployment and integration with existing business processes.
4. Set up a monitoring system to track model performance over time and detect potential drift.