

Lead Scoring Analysis Report

1. Objective

In a competitive business environment, lead conversion plays a critical role in revenue generation. The primary objective of this analysis is to develop a **predictive model** to assess the likelihood of a potential lead converting into a customer. By doing so, the sales team can efficiently allocate resources, focus on high-potential leads, and improve overall conversion rates.

This study leverages **logistic regression** to build a lead scoring model that predicts conversion probability. The insights derived from this model will help the organization make data-driven decisions about customer engagement strategies.

2. Data Preparation & Feature Selection

To ensure the model's robustness, the dataset underwent multiple preprocessing steps before training the logistic regression model.

2.1 Data Splitting

- The dataset was divided into **training (70%)** and **test (30%)** sets to evaluate the model's performance on unseen data.

2.2 Feature Selection

- Recursive Feature Elimination (RFE)** was applied to select the top **17 most relevant variables**.
- Features with **p-value > 0.05** (statistically insignificant) and **VIF > 5** (multicollinearity issues) were eliminated to enhance model interpretability and reduce overfitting.

3. Model Performance & Evaluation Metrics

The trained logistic regression model was tested on the unseen data, and the following results were obtained:

PREDICTED ACTUAL	NOT CONVERTED	CONVERTED
NOT CONVERTED	1583	104
CONVERTED	132	904

3.2 Performance Metrics

- **Accuracy: 91.34%** (Overall correctness of predictions)
- **Precision: 89.26%** (Proportion of predicted conversions that were actually converted)
- **Recall (Sensitivity): 87.26%** (Ability to identify actual conversions)
- **Specificity: 93.84%** (Ability to identify non-converting leads)
- **False Positive Rate (FPR): 6.16%** (Proportion of non-converting leads incorrectly predicted as conversions)

These results indicate that the model performs well in distinguishing between converted and non-converted leads.

4. Key Findings

4.1 Lead Score & Conversion Probability

- The **lead score model shows a conversion rate of 92% on test data** compared to **95% on training data**, indicating that the model is well-generalized and not overfitting.
- An **optimal probability cutoff of 0.45** was chosen based on a trade-off between sensitivity and specificity.

4.2 Important Factors Influencing Lead Conversion

Among all the selected variables, the top three factors that influence the likelihood of conversion are:

1. **Tags_Lost to EINS** – Customers marked as "Lost to EINS" are highly relevant for targeted follow-ups.
2. **Tags_Closed by Horizon** – Leads marked as "Closed by Horizon" indicate strong sales intent.
3. **Lead Quality_Worst** – This factor significantly reduces conversion likelihood, highlighting the need for a better lead qualification process.

These insights help in refining the lead prioritization strategy for maximum efficiency.

5. Business Recommendations

5.1 High Sales Periods (Intern Hiring Phase)

During the **2-month hiring phase**, the company allocates additional interns to the sales team to aggressively push for lead conversion. The following strategy should be implemented:

- Focus on **high-probability leads** (Convert_Probability > 0.27).
- Prioritize phone calls to leads with a high **Lead Score (> 70)**.
- Automate email and SMS follow-ups for medium-probability leads (Lead Score: **40–70**).
- Track and analyze the lead response rate to optimize call timing.

5.2 Low Sales Periods (After Achieving Targets Early)

Once the company reaches its quarterly sales targets, the sales team's focus should shift to efficiency rather than aggressive conversions. The following approach is recommended:

- **Reduce unnecessary phone calls** and only engage with leads having a **Lead Score > 80**.
- Shift focus towards **customer retention, upselling, and exploring new markets**.
- Conduct internal **sales training** and process optimization activities.

These strategic shifts ensure the optimal use of resources based on the company's business cycle.

6. Conclusion

The logistic regression model successfully predicts lead conversion with **high accuracy (91.33%) and precision (89.68%)**. By leveraging **data-driven decision-making**, the organization can:

- ✓ Optimize sales efforts by prioritizing high-potential leads.
- ✓ Improve resource allocation during peak and off-peak periods.
- ✓ Enhance lead nurturing strategies for better conversion rates.

This analysis provides a robust foundation for making **strategic sales decisions** that balance efficiency and effectiveness in the lead conversion process.