Business Analytics

Key Insights and Recommendations

Group 3

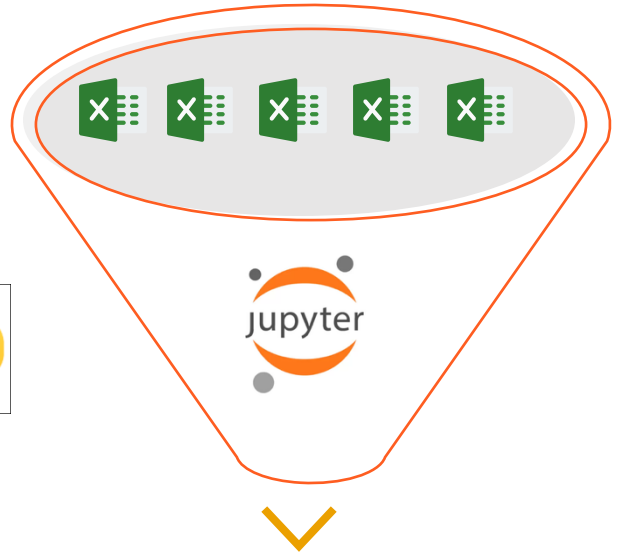
Agenda

1. Key Predictive Features

2. Comparison of Different Modeling Approaches

3. Conclusion

Data Sources



Engineer features

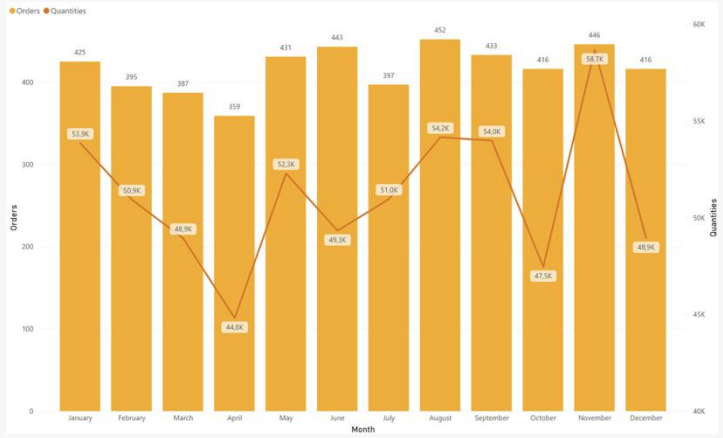
Data is already cleaned

Star scheme from PowerBI

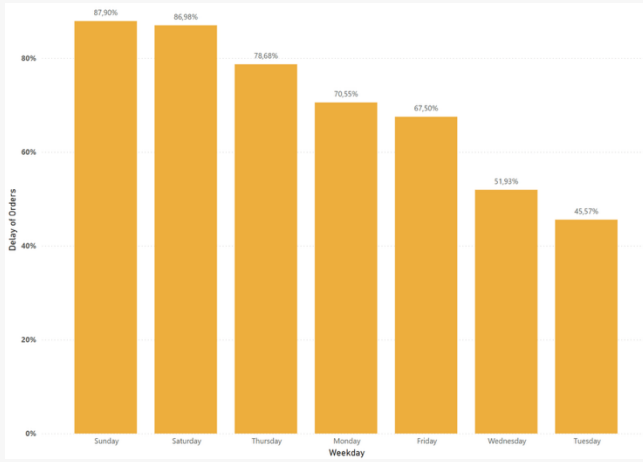
Relations between "**transactions**" and:

- Material_code ↔ item_code/ **material**
- Customer ↔ name/ **customer**
- Logistic_partner_code ↔ id/ **logistic partner**
- Warehouse_code ↔ id/ **warehouse**

Recap - Key Predictive Features



Orders and quantities



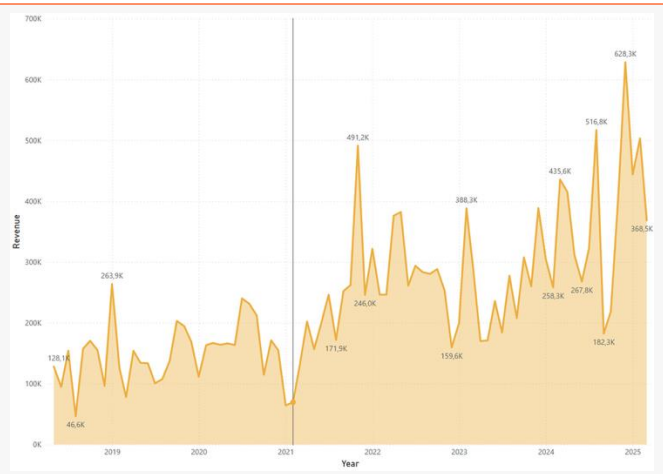
Delay of Orders by Weekday

KPI

2,76
Days Late

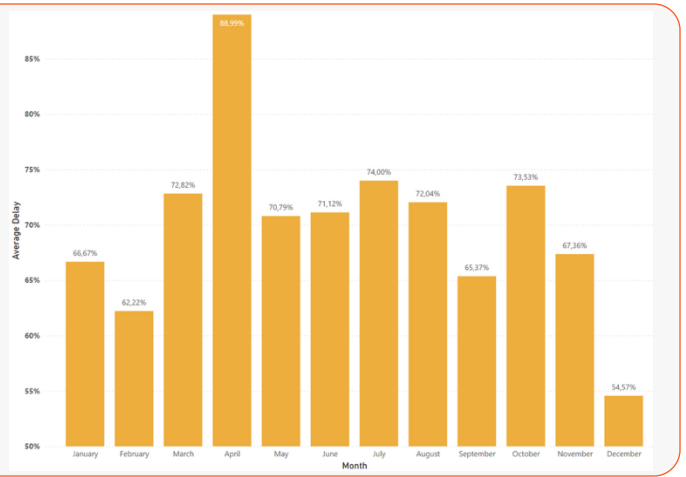
70%
Late Orders

5000
Order



Revenue

Average Delay by Month



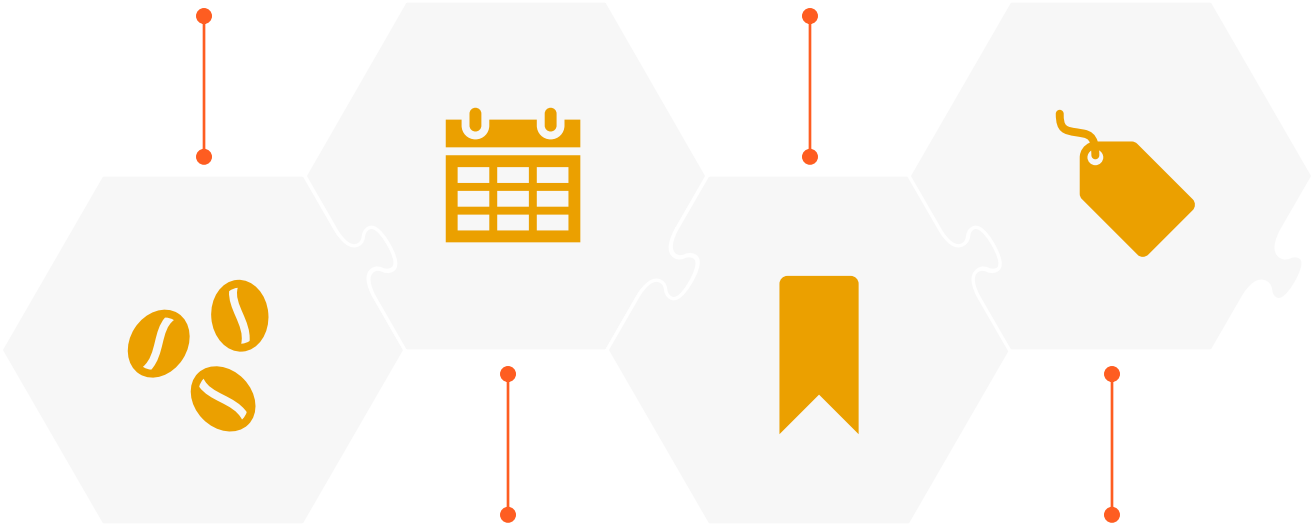
Feature Selection

quantity

in kilograms per order

weekday

When was it delivered



months

When was it ordered

Price category

Price range of each order

```
features = (['quantity'] +
[col for col in transactions3.columns if col.startswith('Months_')] +
[col for col in transactions3.columns if col.startswith('Weekday_')] +
[col for col in transactions3.columns if col.startswith('price_category_')]);
```

```
['quantity',
 'Months_April',
 'Months_August',
 'Months_December',
 'Months_February',
 'Months_January',
 'Months_July',
 'Months_June',
 'Months_March',
 'Months_May',
 'Months_November',
 'Months_October',
 'Months_September',
 'Weekday_Friday',
 'Weekday_Monday',
 'Weekday_Saturday',
 'Weekday_Sunday',
 'Weekday_Thursday',
 'Weekday_Tuesday',
 'Weekday_Wednesday',
 'price_category_0-20',
 'price_category_20-40',
 'price_category_40-60',
 'price_category_60-80',
 'price_category_80-100',
 'price_category_100-120',
 'price_category_120-140',
 'price_category_140-160',
 'price_category_160-180',
 'price_category_180-200',
 'price_category_200-220']
```

Comparison of Different Modeling Approaches

The best results ?

Features :

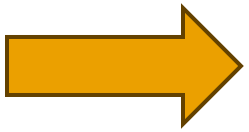
Quantity

Months

Weeks

Price

Depth : 80



Accuracy: 99.98 %
Precision: 100.00 %
Sensitivity: 99.97 %

Typical case of Overfitting !!!

Management Summary

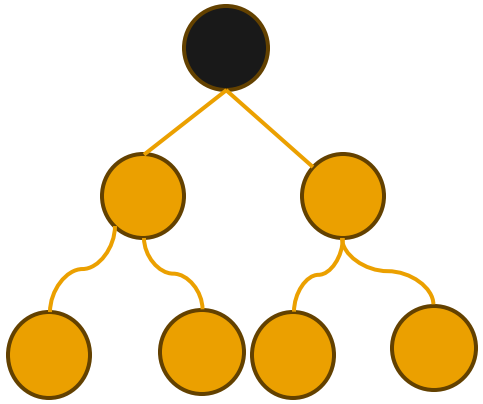
Features :

Quantity

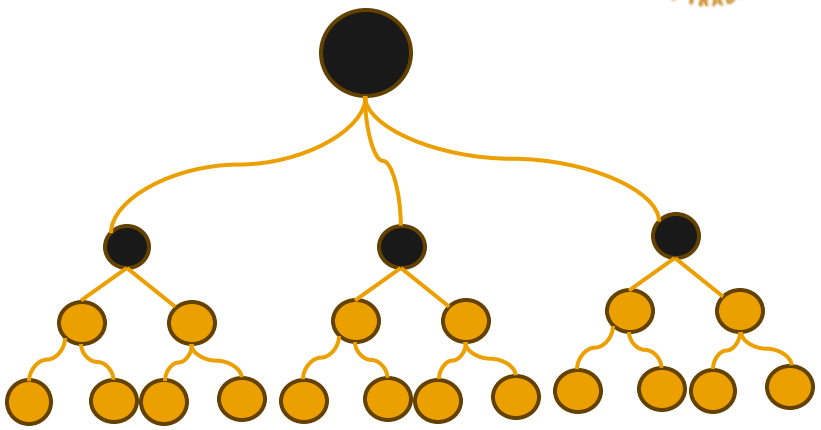
Months

Weeks

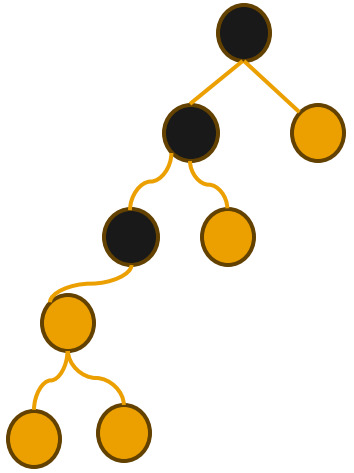
Price Category



Decision Tree

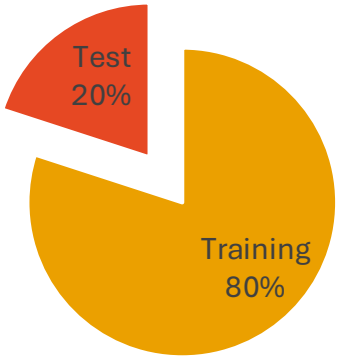


Random Forest

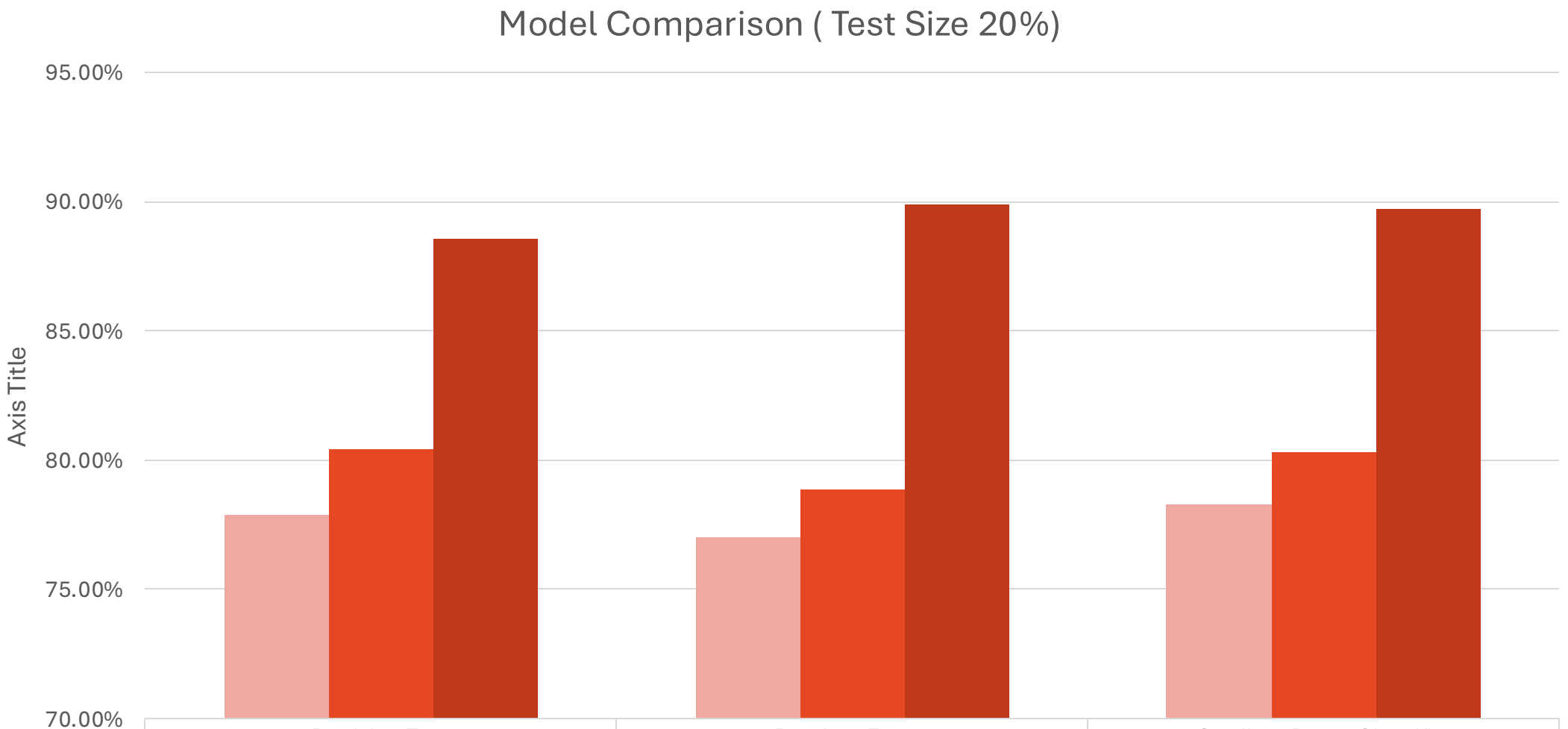


Gradiant Boost Classifier

Split

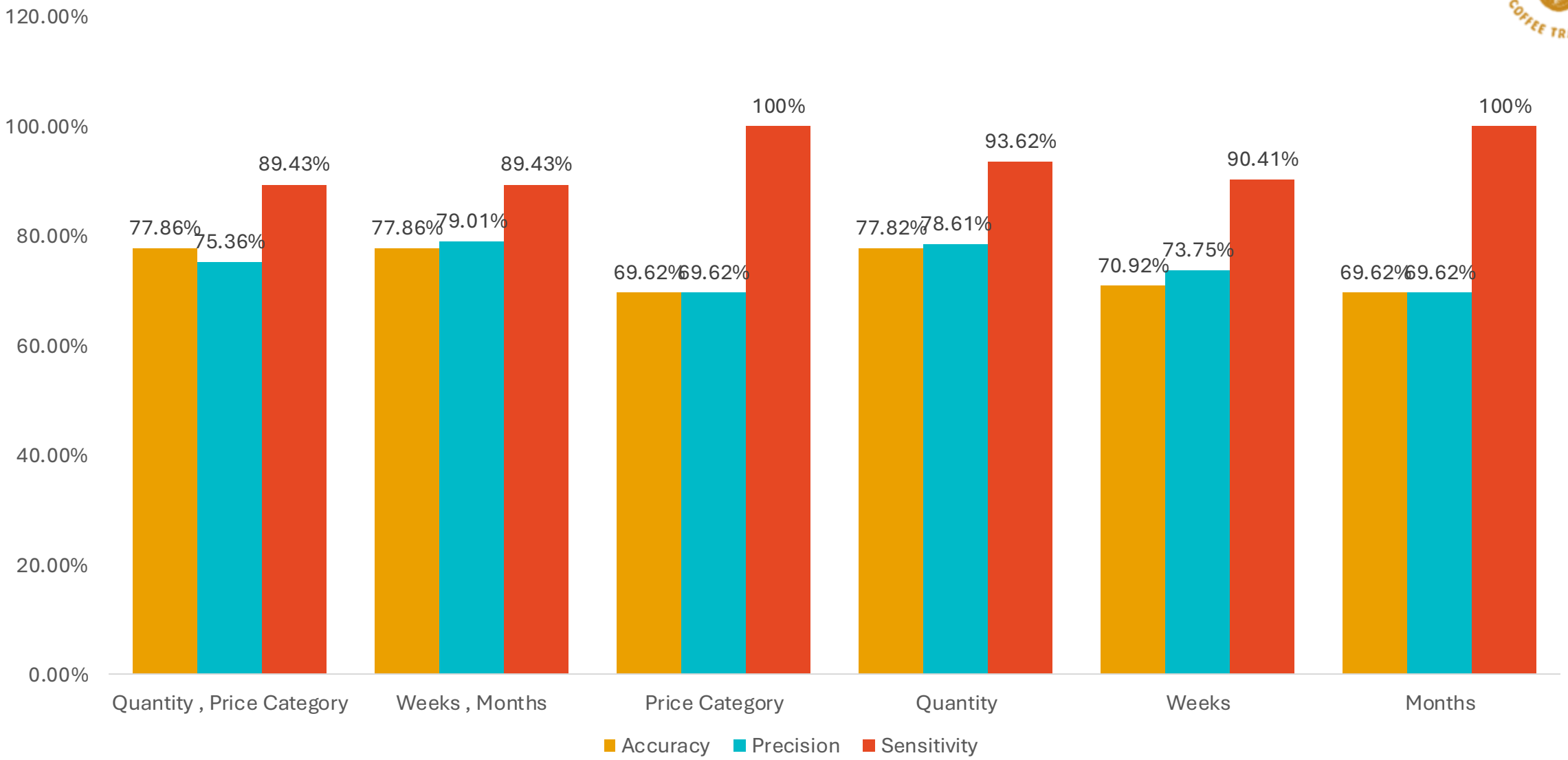


Model Comparison

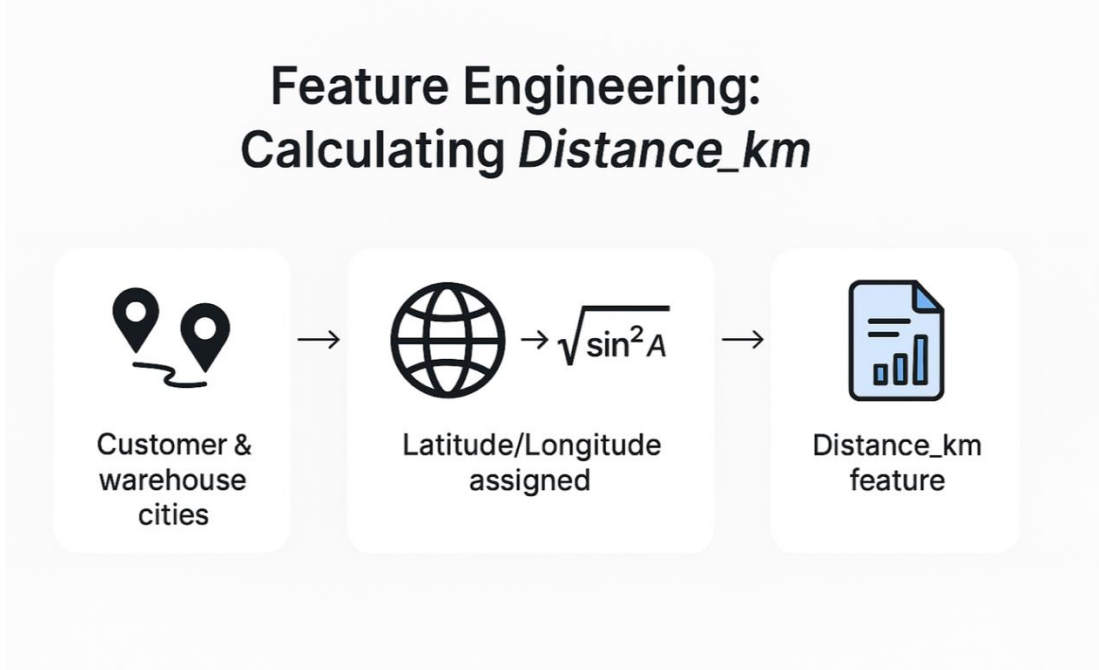


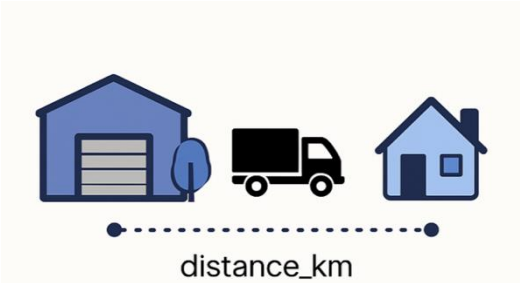
	Decision Tree	Random Forest	Gradiant Boost Classifier
Accuracy	77.90%	77.00%	78.30%
Precision	80.45%	78.88%	80.32%
Sensitivity	88.56%	89.90%	89.75%

Decision Tree

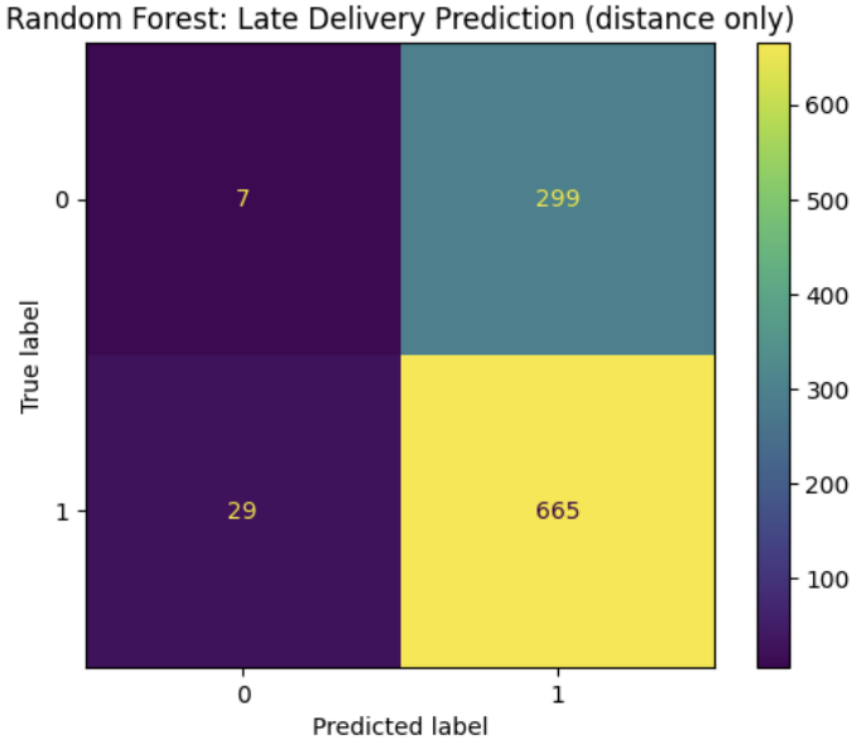


Feature : Distance with Random Forest





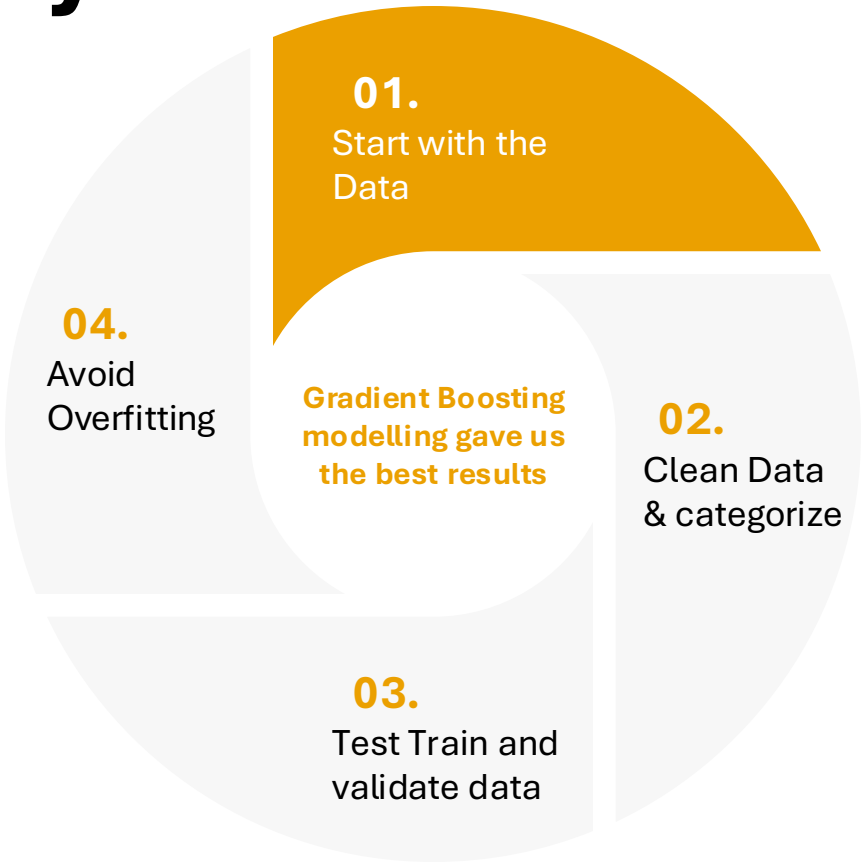

```
[Customer City] + [Warehouse City]
        |
   (assign coordinates)
        |
   [Haversine Formula]
        |
   [distance_km Feature]
        |
   [Random Forest Classifier]
```

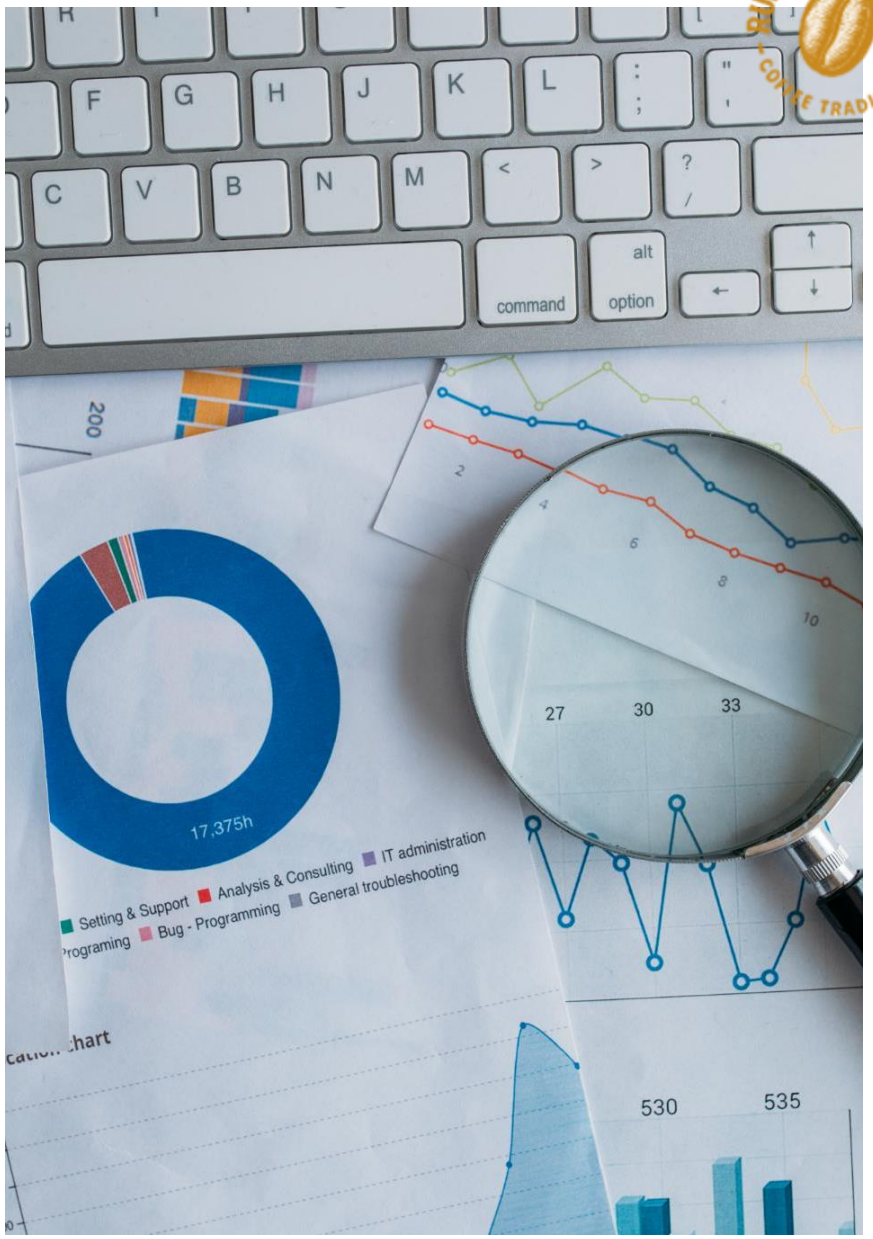


	precision	recall	f1-score	support
0	0.19	0.02	0.04	306
1	0.69	0.96	0.80	694
accuracy			0.67	1000
macro avg	0.44	0.49	0.42	1000
weighted avg	0.54	0.67	0.57	1000

Conclusion

Final Thoughts and Takeaways





Thank you for your attention

Ananta Das, 23595221

Timo Gottsche, 22801199

Wassilissa Golowatjuk, 23616466

Mario Sicaja, 23595293

Stefan Lößner, 23595260