Enhancing Credit Analysis & Assessment using Geospatial Techniques

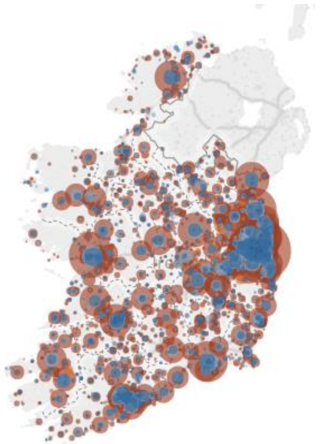
By:

Deepak Kumar Gupta
Shruti Goyal

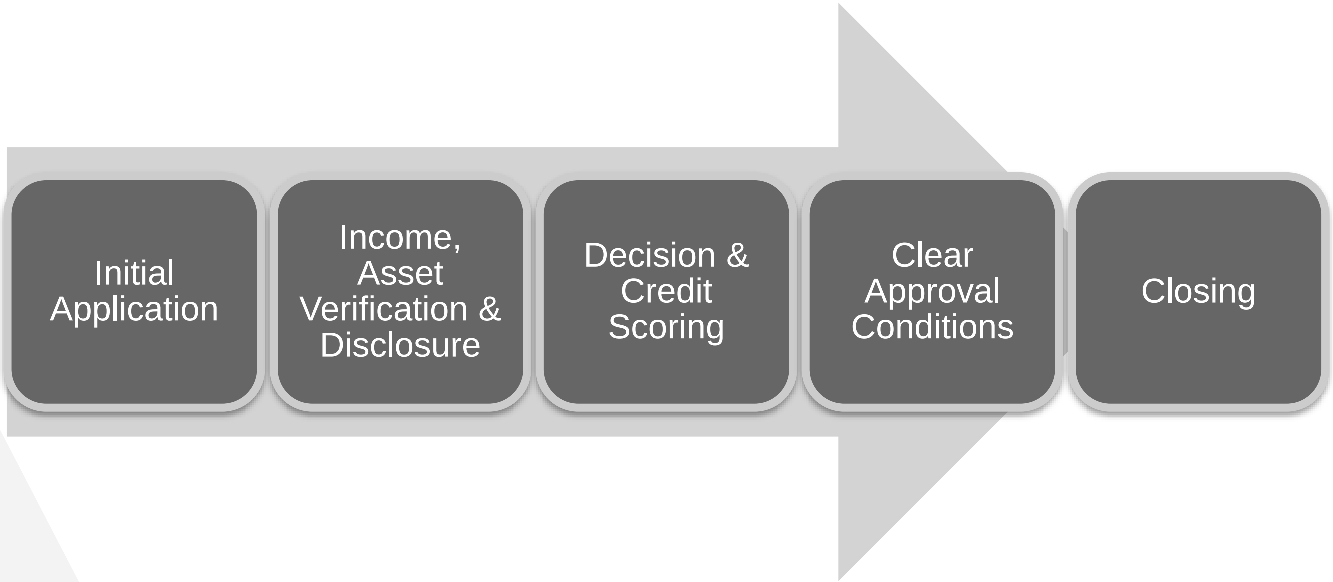
Supervisors:

Prof. Peter Keenan, UCD
Selwyn Hearns, KPMG

Business Problem



Loan Process



Previous Work

Previous Work

Method	Authors
Linear Regression	Lee and Chen (2005); Hand and Henley (1997)
Discriminant Analysis	Fisher (1936); Durand et al. (1941); Altman (1968); Eisenbeis (1978); Zhou et al. (2016); Liberati et al. (2017)
Logistic Regression	Hosmer et al. (1989); Altland (1999); Nie et al. (2011); Abdou et al. (2008); Bensic et al. (2005); Joanes (1993)
Decision trees	Kohavi and Quinlan (2002); Breiman et al. (1984); Zhang et al. (2010); Zekic-Susac et al. (2004); Zhou et al. (2008); Huang et al. (2007); Xia et al. (2017); Koh et al. (2015); Koutanaei et al. (2015)
Neural networks	Demuth et al. (2008); West (2000); Gately (1995); Presky et al. (1996); Ghosh and Reilly (1994); Desai et al. (1996)

Previous Work

- Durand et al. (1941) used the Discriminant analysis for modelling a scoring system that gives a prediction about loan repayment.

- Limitations:
  - Reduction in dimensionality
  - Improper estimation of classification error
  - Using linear functions instead of quadratic function

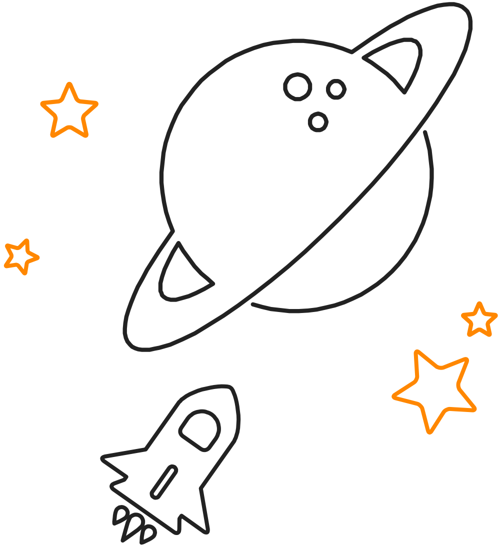
Previous Work

- Huang et al. (2007) used decision tree along with support vector machines to build credit scoring model

- Limitations of SVMs:
  - Long training time
  - Black box nature of model(similar to NNs)

Previous Work

- Carling and Lundberg (2005) combined the geographical information with loan data to examine the credit rationing.
- Finding:
  - Technological changes out weight needs of geographical proximity
  - Increasing distance between lender and borrower might be alarming



Methodology

Data

Data Sources:

- Credit Rating and Loan portfolio from Bank
- Property price register
- Employment Rate(CSO)

Key Variables:

- DefaultedLoans
- CreditRating
- PropertyValue
- LoanBalance
- LTV
- InterestType
- ProbationaryLoans
- County
- Town
- AddressLatitude
- AddressLongitude

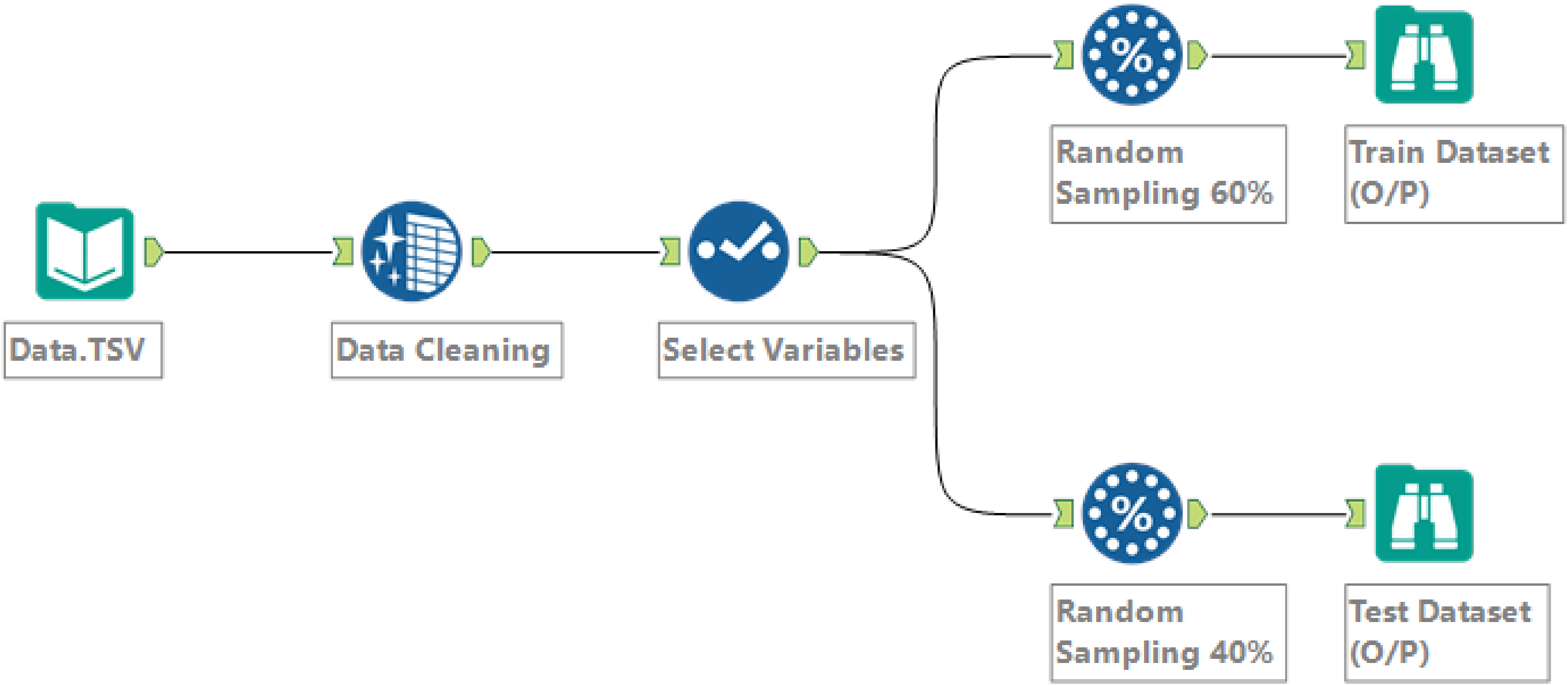
Data Assumptions

- Geospatial data such as address latitude and address longitude is assumed to depict geospatial location of a property correctly
- Dimensions of house and size of the house(number of rooms, bathrooms, lawn, etc.) are not considered during model development
- Factors such as neighbourhood, amenities etc. are not considered that affects the property price in the market

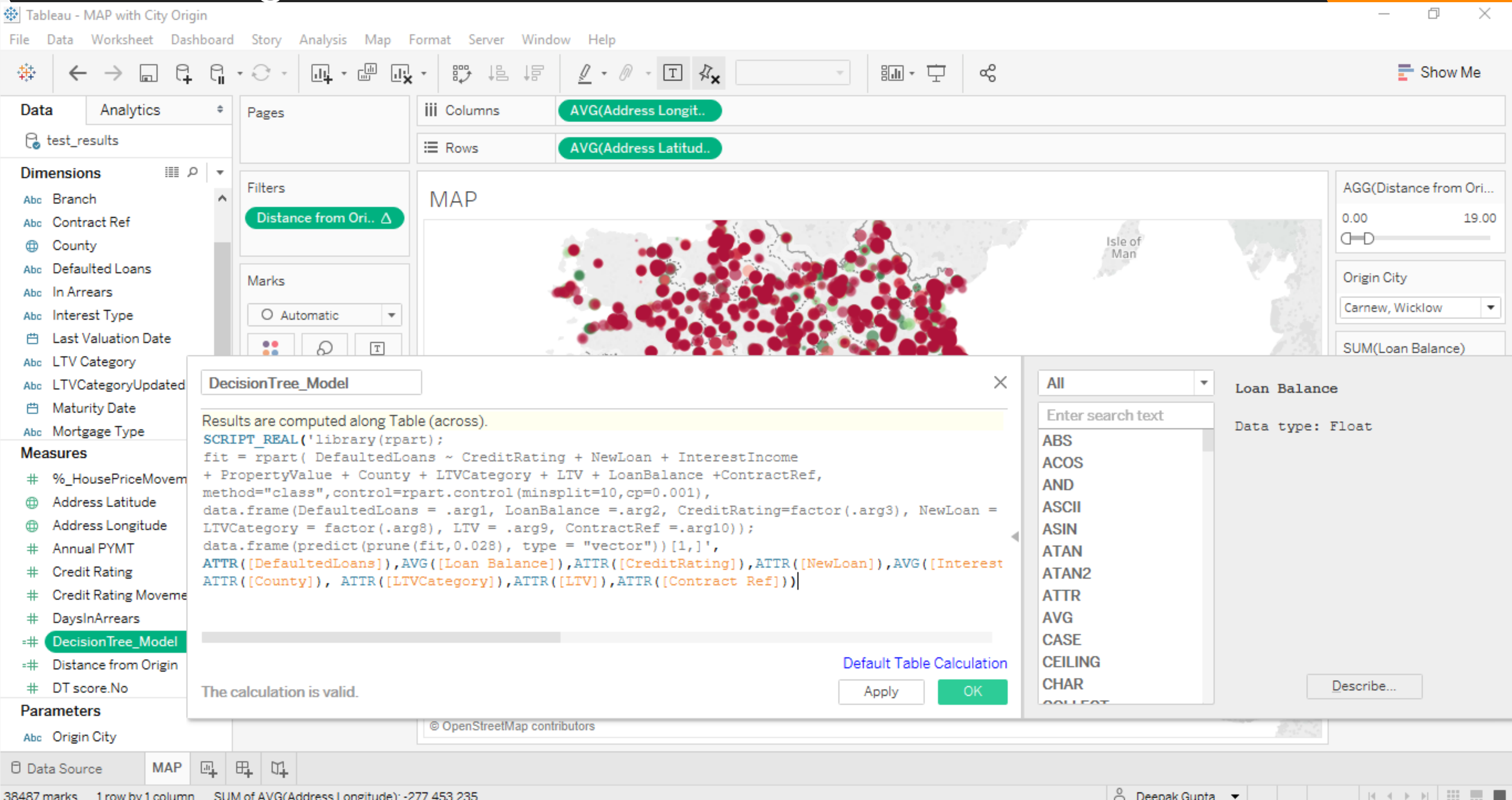
Algorithms Considered

- Logistic Regression
- Decision Tree(CART)

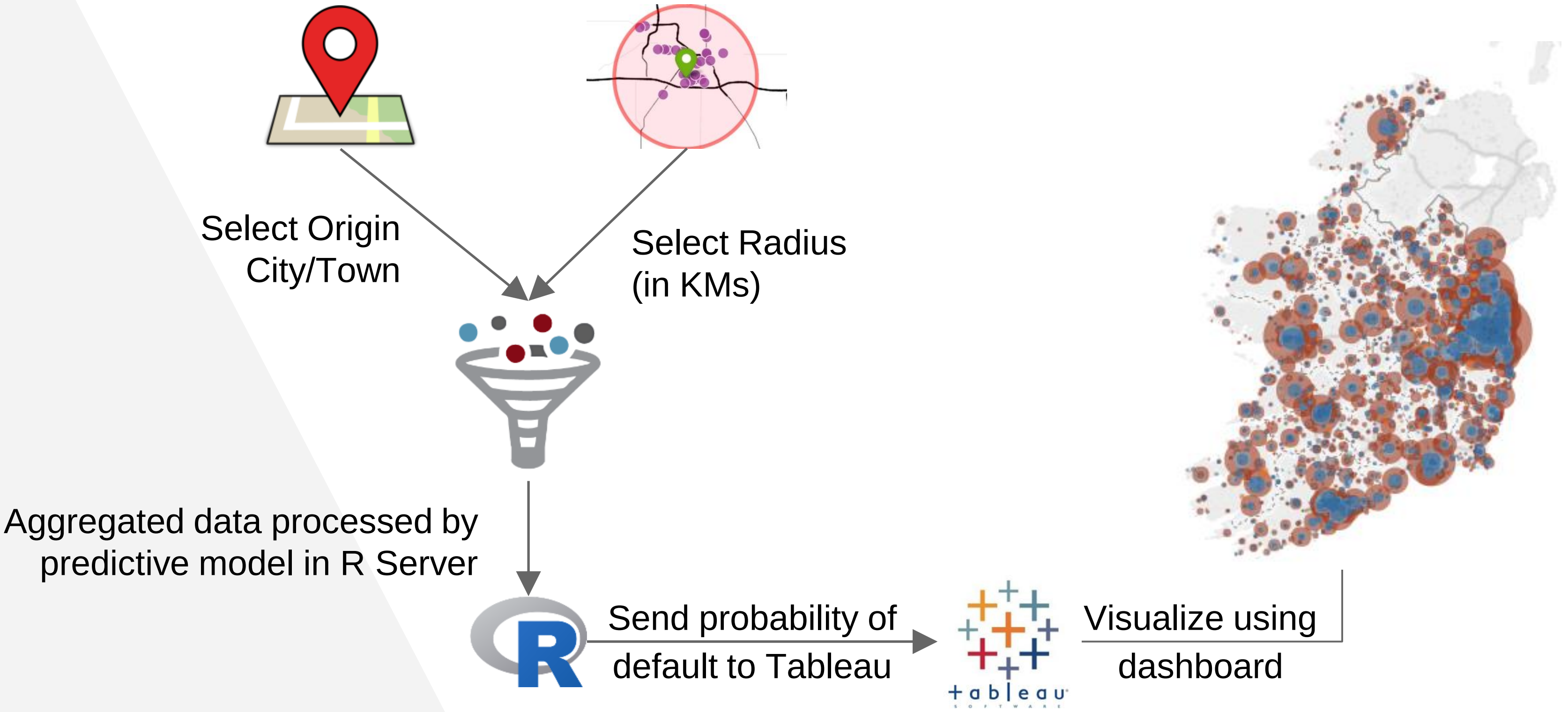
Data Processing

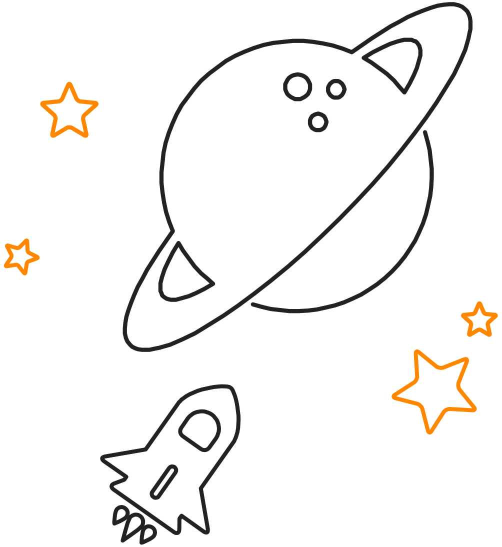


Connecting Tableau with R



Our Approach





Analysis

Distribution of Loan Accounts

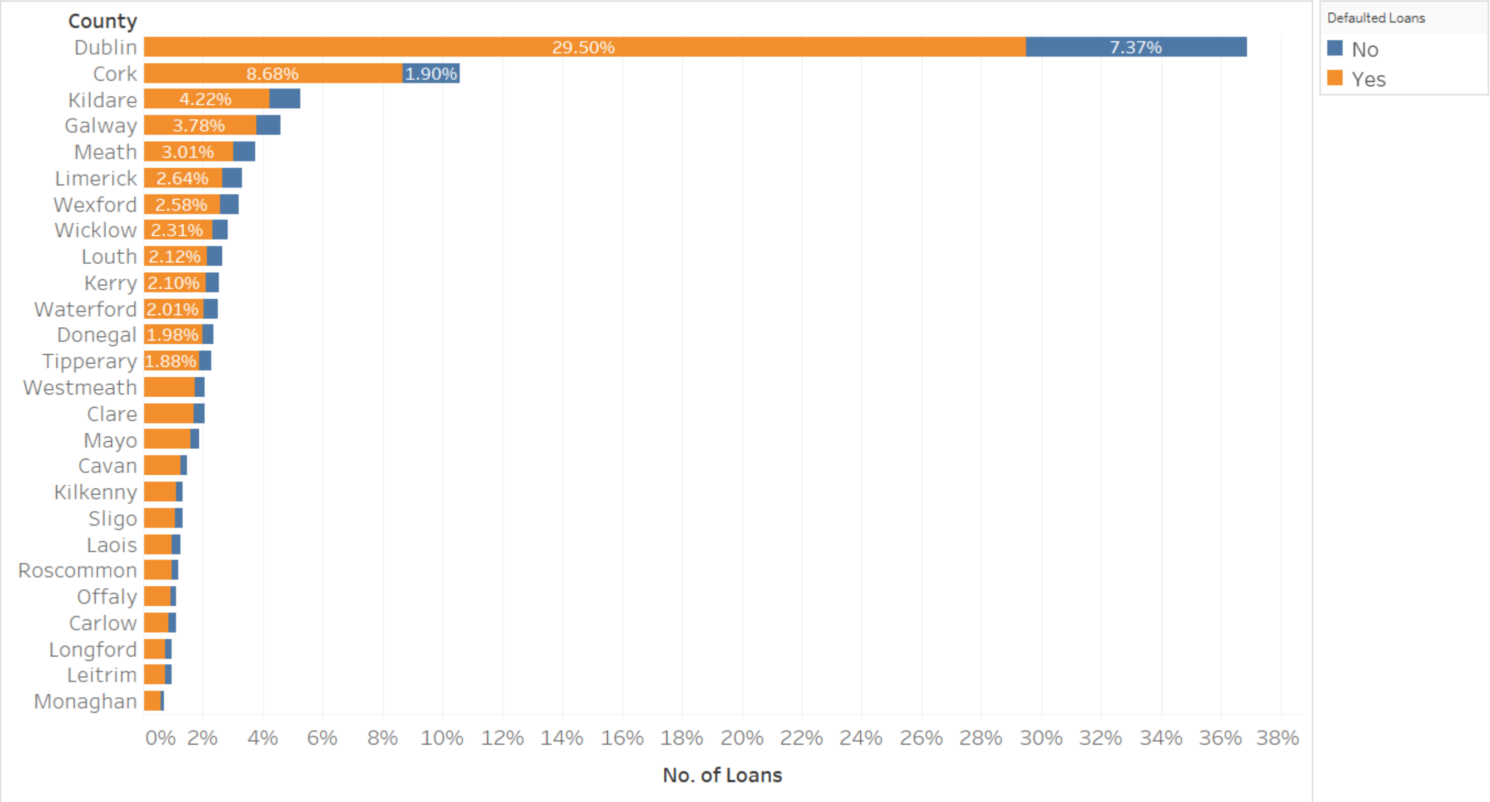
Original Data

	Defaulted Loans	
Credit Rating	No	Yes
1	20.33%	
2	20.87%	0.03%
3	17.00%	0.03%
4	13.45%	
5	1.09%	27.19%

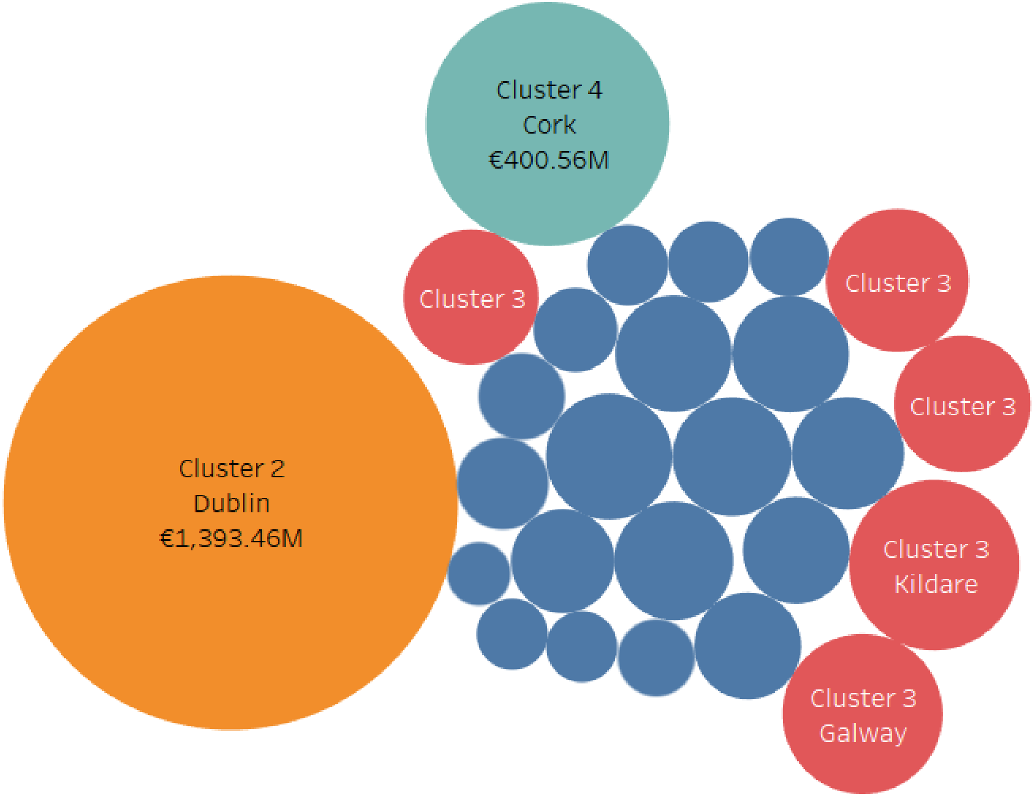
Normalized Data

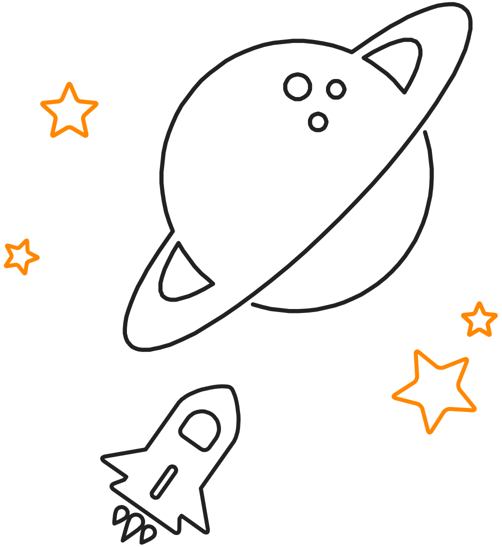
	Defaulted Loans	
Credit Rating	No	Yes
1	1.31%	19.02%
2	0.69%	20.21%
3	7.24%	9.79%
4	8.01%	5.44%
5	1.86%	26.42%

Distribution of Loan Accounts vs County



Clustering County based on Loan Balance





Results

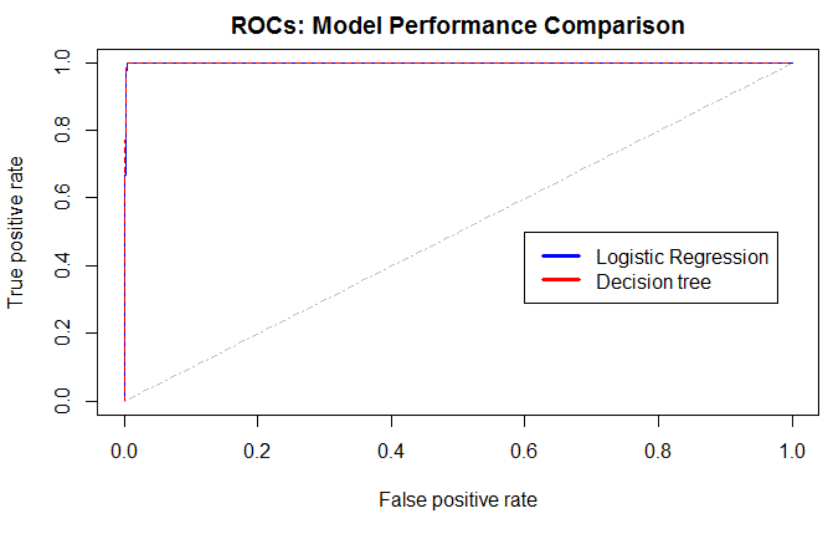
Logistic Regression and Decision Tree

Model Performance Metrics

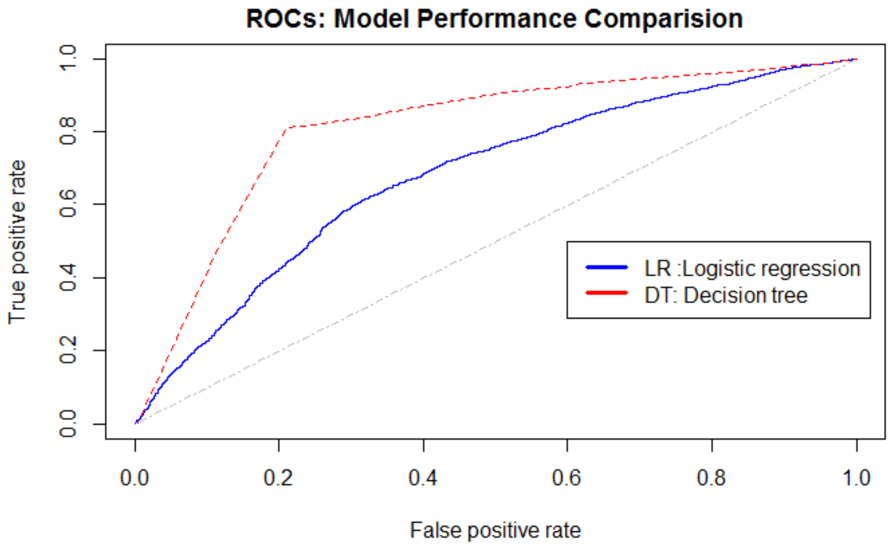
	Logistic Regression			Decision Tree		
Data/Measure	**AUROC**	**KS**	**Gini**	**AUROC**	**KS**	**Gini**
Original Data	**99.82**	**15**	**10**	**99.72**	**99.38**	**99.44**
Normalized Data	**67.61**	**24**	**16**	**81.4**	**59.96**	**62.8**

Model Performance Comparison

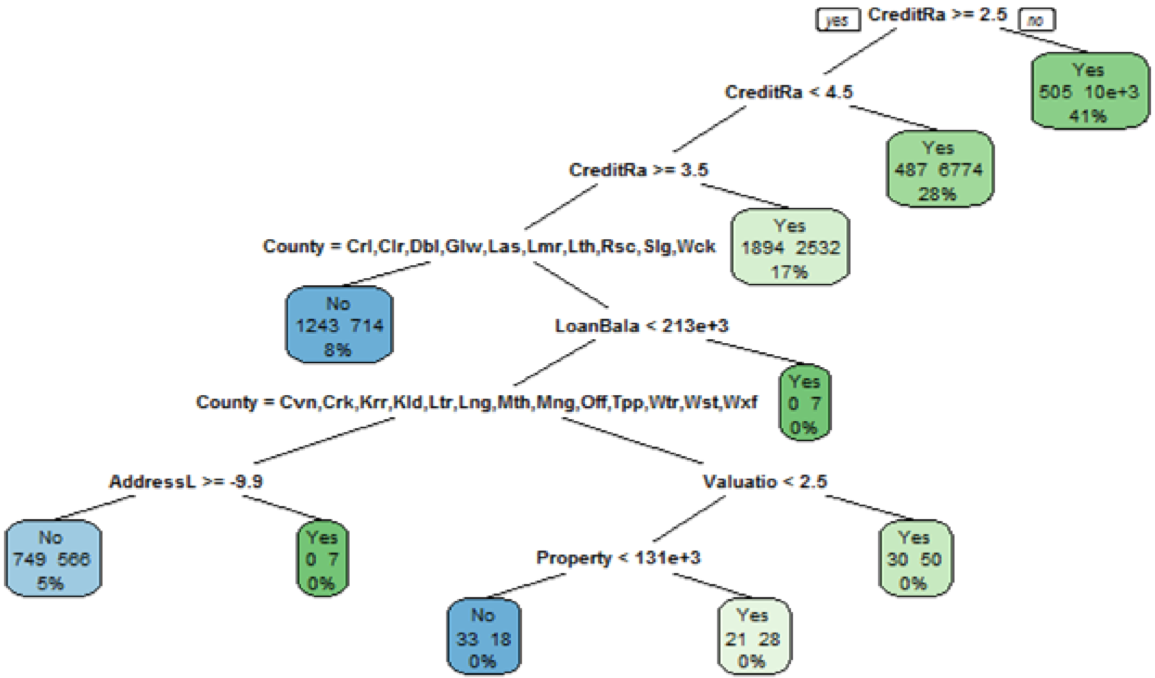
Original Data



Normalized Data



Decision Tree +



+ Limited number of nodes in tree due to page size constraint

[Dashboard Link]

Measure of Success

47%

Auditors feel this dashboard is extremely useful

70% users

Feel that geospatial techniques have enhanced credit assessment

60% users

Likely to recommend this dashboard to colleagues

End Users Profile+

Selwyn H
Head of IRM Audit at KPMG Ireland

James F
Auditor, KPMG

Simon M
Head of Data Science, Segmatic

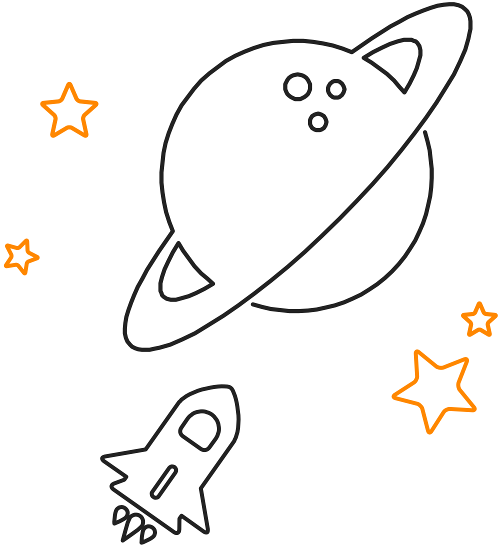
John M
Software Developer, Kinesene Ltd

Ciaran F
Senior Product Consultant, Tableau

John L
Banking Director, SAS UKI

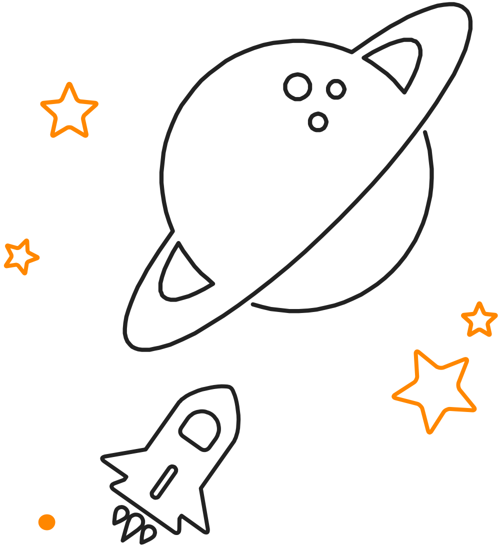
+ selective users profile out of 15

Business Contributions



Business Contributions

- Interactive way to identify patterns in datasets to drill down into problem areas
- Well timed potential issues indicators that adhere to provisions of audit processes and assessment of residential loans
- Better and greater coverage of problem areas and increased focus on judgemental loan applications
- Integration of useful and relevant market data and economic indicators for enhanced loan assessment



Recommendations

Recommendations

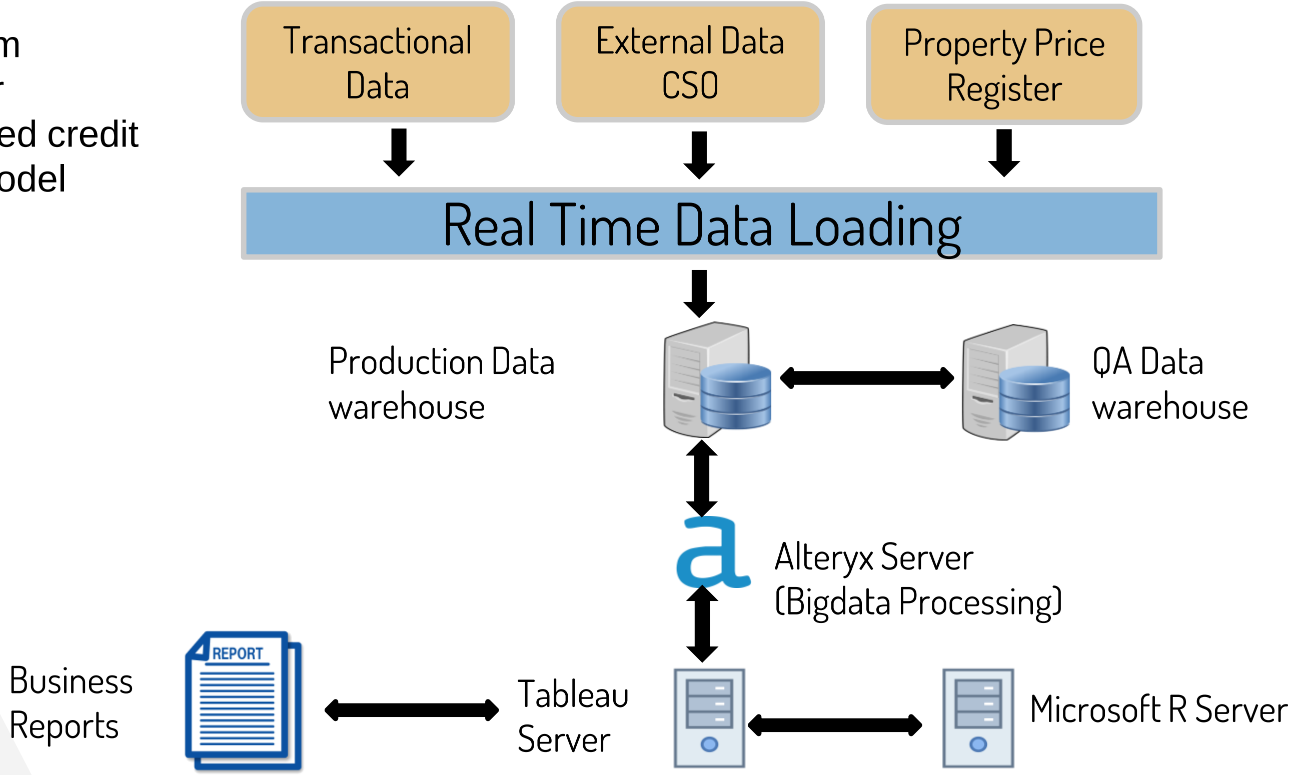
Dashboard
Real transactional data can improve the performance of Tableau dashboard which acts as decision support system

Geospatial Data
Credit assessment can be enhanced if it includes information such as house coordinates

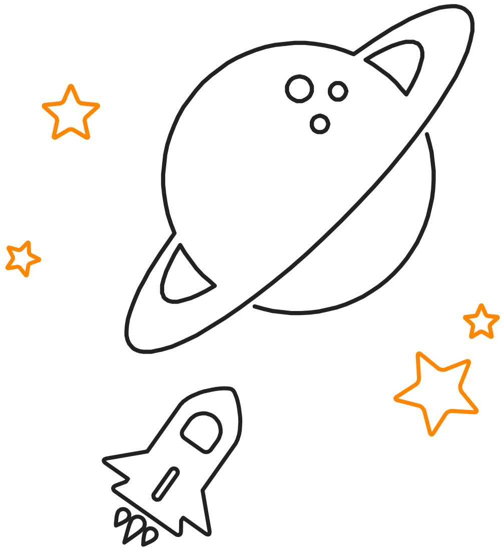
External Factor
The model can perform better when trained with large number external factors such medical information, average salary in neighbourhood

Recommendations

Detailed system architecture for geospatial based credit assessment model

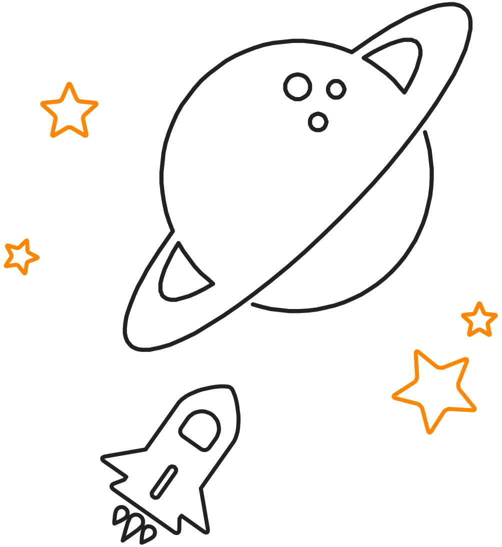


Our Learning



Our Learning

- Importance of Data Processing
- Deployment of RServer
- How performance of a model can vary with data
- Human interaction is must in credit analysis



Conclusion

UCD Michael Smurfit Graduate Business School

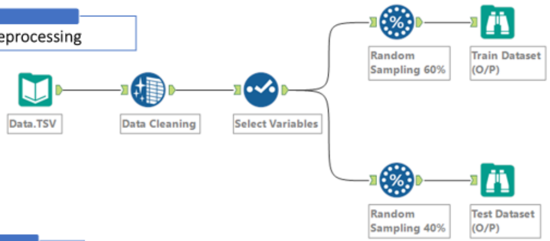
Enhancing Credit Analysis & Assessment Using Geo-Spatial Techniques

Deepak Kumar Gupta, Shruti Goyal, Supervisor: Prof. Peter Keenan
MSc Business Analytics (Practicum)

Problem Statement: To improve information and underlying data with regard to loan performance, particularly in the property sector using Geo-spatial techniques with interactive dashboard.

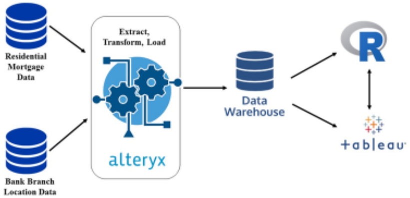
Phase1

- Data Preprocessing



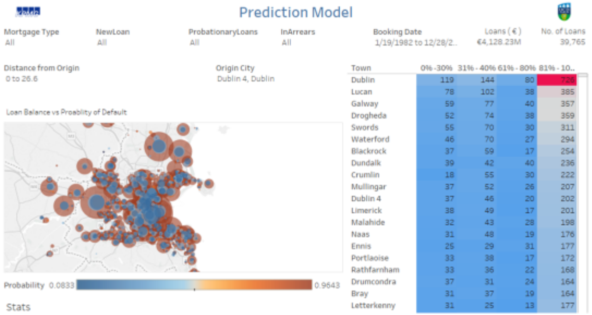
Phase 2

- Data Model



Phase 3

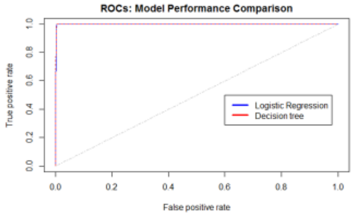
- Visualization & Dashboard



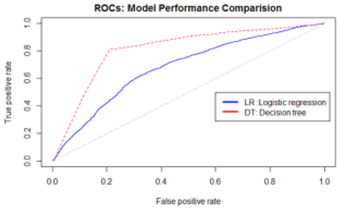
Results:

Model performance evaluated with original data and normalized data on Logistic regression and decision tree. Both model gave accuracy of 99.89% but with normalized dataset, decision tree outperformed logistics regression by 20% better accuracy.

Model	Logistic Regression			Decision Tree		
Data/Measure	AUROC	KS	Gini	AUROC	KS	Gini
Original Data	99.82	99.25	99.64	99.72	99.38	99.44
Normalized Data	67.61	28.68	35.22	81.4	59.96	62.8



A. Original Data



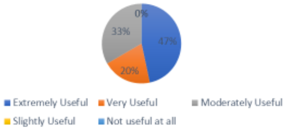
B. Normalized Data

Business Contributions:

1. Auditors and credit analysts can use the dashboard to identify potential default loans
2. Financial institutions can develop new business strategies to offer loans targeted to particular areas

Conclusion:

As an auditor, How useful do you think this dahsboard is?



- 47% user said this dashboard is extremely useful as auditing tool for loan portfolios.
- Decision tree perform better over logistic regression model
- Algorithm performance can be improved when real transactional data plugs in.