

Motor Vehicle Accident Analysis

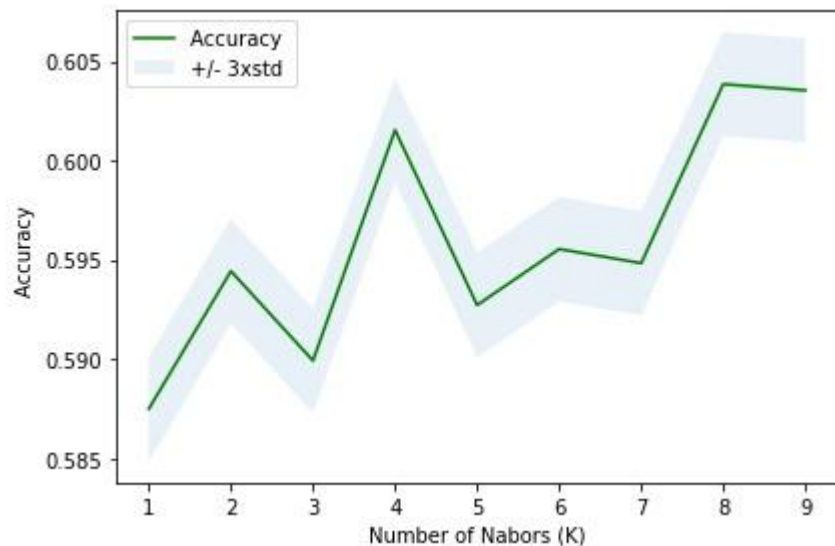
Applied Data Science Capstone

Charles Gaglac

The Problem

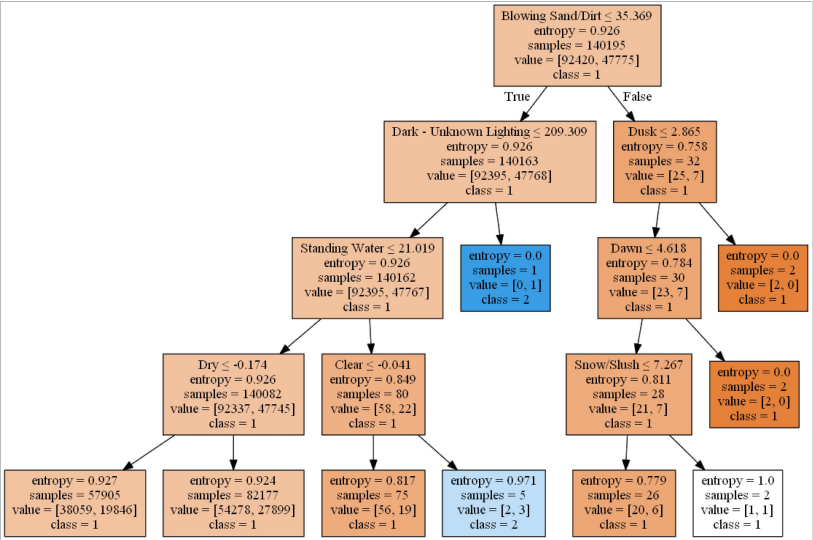


k-Nearest Neighbor



	precision	recall	f1-score	support
1	0.66	0.82	0.73	23128
2	0.35	0.19	0.24	11921
accuracy			0.60	35049
macro avg	0.50	0.50	0.49	35049
weighted avg	0.55	0.60	0.57	35049

Decision Tree



	precision	recall	f1-score	support
1	0.66	1.00	0.80	23128
2	1.00	0.00	0.00	11921
accuracy			0.66	35049
macro avg	0.83	0.50	0.40	35049
weighted avg	0.78	0.66	0.52	35049

Support Vector Machine

	precision	recall	f1-score	support
1	0.66	0.83	0.73	23128
2	0.34	0.17	0.23	11921
accuracy			0.60	35049
macro avg	0.50	0.50	0.48	35049
weighted avg	0.55	0.60	0.56	35049

Logistic Regression

	precision	recall	f1-score	support
1	0.66	1.00	0.80	23128
2	0.00	0.00	0.00	11921
accuracy			0.66	35049
macro avg	0.33	0.50	0.40	35049
weighted avg	0.44	0.66	0.52	35049

Key Takeaways

- Road conditions and weather are linked to motor vehicle accidents
- Decision Tree and Logistic Regression are the best models
- Dataset imbalance cripples the classification of personal injury outcomes