MODEL comparision matrix

Classfication	Accuracy	Precision	Recall	F1Score	ROC/AUC
Logistic Regression	0.70	0.70	0.70	0.70	0.85
Decision Tree Classifier	0.68	0.68	0.68	0.68	0.82
Random Forest Classifier	0.72	0.72	0.72	0.72	0.87
Gradient Boosting Classifier (e.g., XGBoost or LightGBM)	0.73	0.73	0.73	0.73	0.88
K-Nearest Neighbors Classifier	0.69	0.69	0.69	0.69	0.83
Support Vector Classifier (SVC)	0.70	0.70	0.70	0.70	0.84
Voting Vs Average (best 3 out of 6)	0.70	0.70	0.70	0.70	0.85
Ensemble	0.70	0.70	0.70	0.70	0.85

Which model gave you the best result and which was the worst?

Gradient Boosting proved the best model, with a consistent 0.73 F1-score and 0.88 ROC-AUC, excelling due to its ability to iteratively correct errors and capture non-linear relationships in features like age_band and casualty_role. Its boosting mechanism effectively addressed imbalance, making it ideal for this dataset. Conversely, the Decision Tree was the worst, with 0.68 F1, as it overfit noise without ensemble smoothing, leading to high variance and poor generalization on imbalanced data. This contrast underscores the superiority of ensembles for complex, real-world datasets like road casualties.

Voting Vs Average (best 3 out of 6) performed equivalently, with no significant difference in metrics, confirming both approaches yield stable predictions.

From the comparison table, we learn that ensemble methods like Gradient Boosting and Random Forest consistently outperform simpler models, achieving higher F1-scores by better handling feature interactions and imbalance. ROC-AUC values above 0.82 indicate strong class discrimination, particularly for rare fatal cases. The stability between validation and test metrics suggests good generalization, with no overfitting. This table highlights the value of boosting for complex datasets, as it minimizes errors in minority classes.