

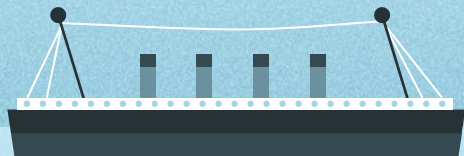
Titanic Survival Prediction

Machine Learning Project
Group – 59

Aakash Sagar Sreerala – A20547846

Vaishnavi Bhaskara – A20546861

Pavani Reddy Podduturi – A20545675



— Problem Statement —

- Accurately predict Titanic passenger survival using machine learning amidst historical data complexities.
- Develop a robust predictive model to differentiate survivors from non-survivors based on all the data that we have about the passengers.
- Explore machine learning's utility in understanding historical events like the Titanic disaster.

— Methodology —



1. Data Collection: Acquire the Titanic dataset containing passenger attributes like age, gender, class, and ticket details.

2. Data Preprocessing: Cleanse and prepare the dataset by handling missing values, encoding categorical variables, and scaling numerical features.

3. Model Selection: Explore and select appropriate classification algorithms such as Logistic Regression, Decision Trees, Random Forest, and Support Vector Machines.

4. Model Training and Evaluation: Train each selected model on a subset of the dataset and assess their performance using metrics like accuracy, precision, recall, and F1-score.

5. Model Optimization and Deployment: Fine-tune the hyperparameters of the chosen model(s) and deploy the optimized model(s) to make predictions on new data for real-world application.

— Data Collection —

We have collected the dataset from kaggle

[2]:

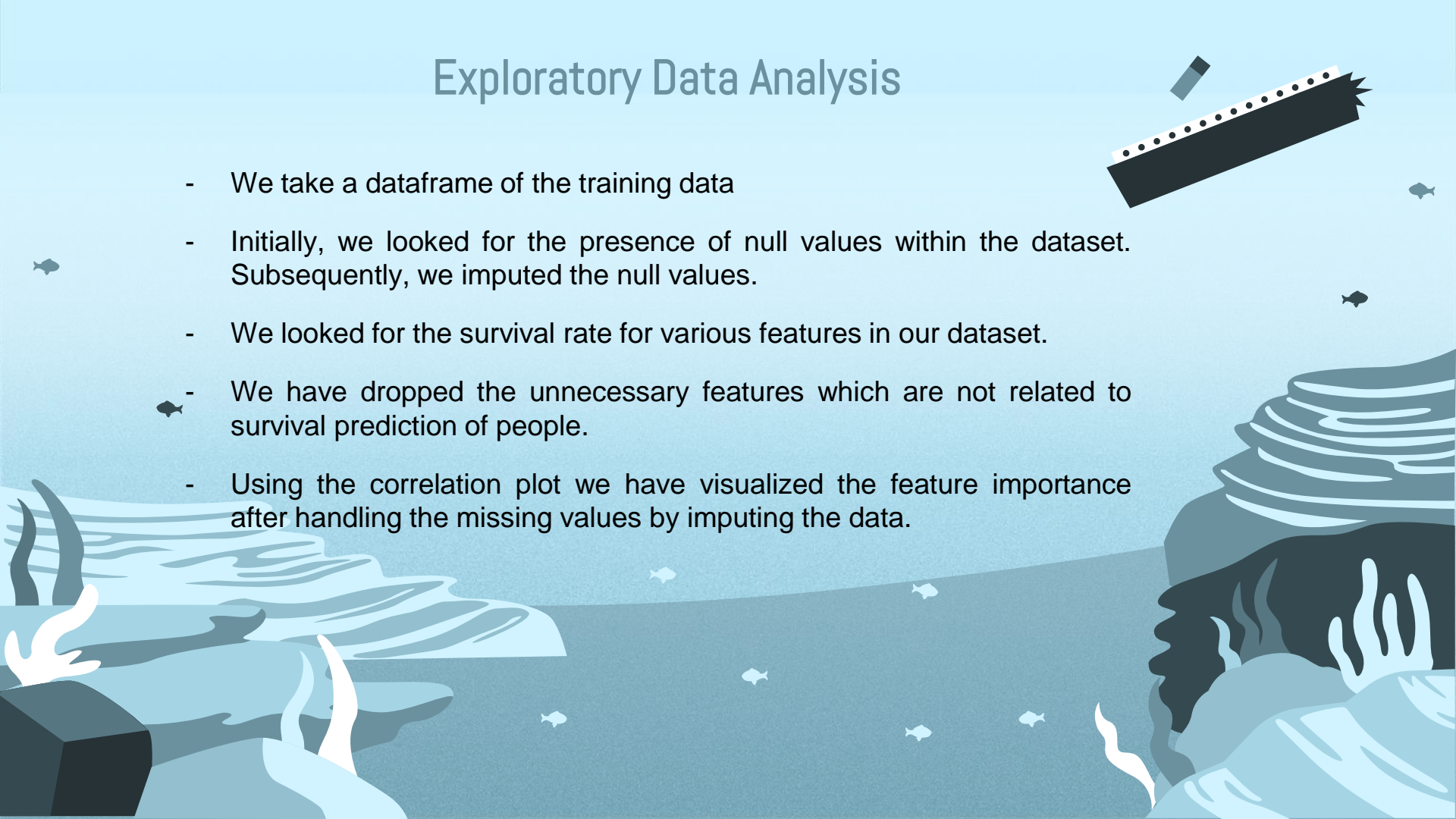
	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	C
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN	S

[3]:

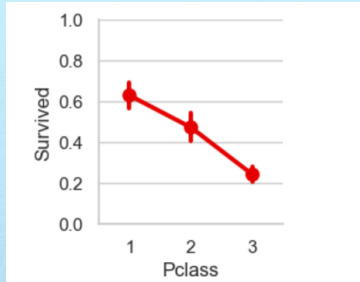
	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	13.00	NaN	S
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.00	B42	S
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	23.45	NaN	S
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30.00	C148	C
890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376	7.75	NaN	Q

Exploratory Data Analysis

- We take a dataframe of the training data
- Initially, we looked for the presence of null values within the dataset. Subsequently, we imputed the null values.
- We looked for the survival rate for various features in our dataset.
- We have dropped the unnecessary features which are not related to survival prediction of people.
- Using the correlation plot we have visualized the feature importance after handling the missing values by imputing the data.

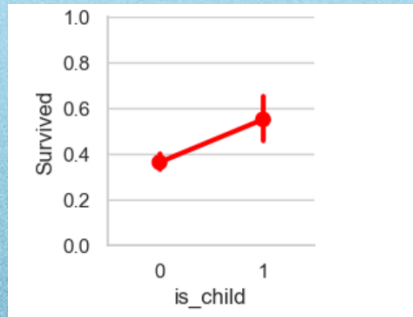
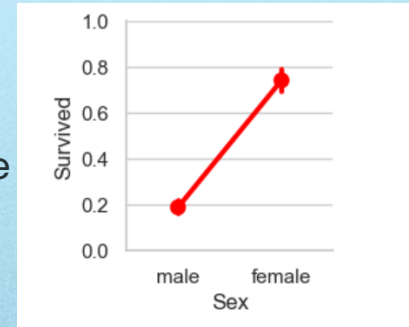


Pictorial Representation of Survival Rate



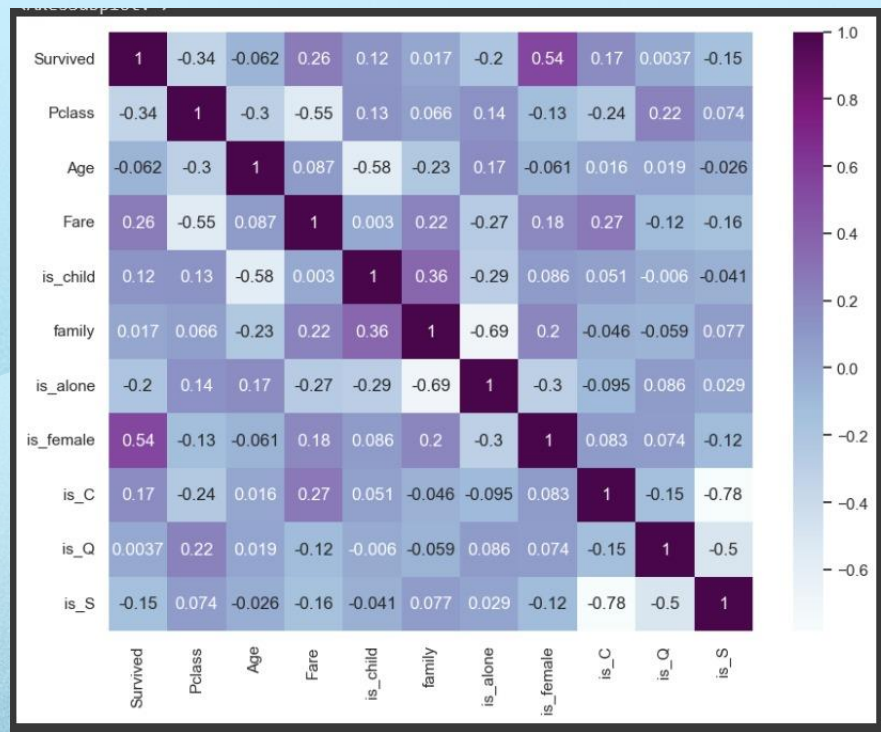
Passenger Class vs Survival Rate

Gender vs Survival Rate



Age vs Survival Rate

Correlation plot



—Model Training—

1.

We have used 80% of data for training and 20% of the data for testing.

3.

Then we have used Random Forest Classifier which gave us an accuracy of 81.01% on test data.

2.

We have checked the accuracies of different models like SGDClassifier, GradientBoostingClassifier, RandomForestClassifier, Logistic Regression, SVCClassifier and Confusion Matrix

Results

S.no	Models Tested	Accuracy
1	SGDClassifier	75.98
2	GradientBoostingClassifier	78.77
3	RandomForestClassifier	81.01
4	Logistic Regression	74.86
5	SVCClassifier	69.83
6	Confusion Matrix	80.45

Finally, we have used Random Forest Classifier which gave us an accuracy of 81.01% on test data.

THANK YOU

