Spec2Model Submission Sheet

Team Information -

- **Team Name: Endeavour**

- **Problem Chosen:** A

- **Member 1:** Ankit Kumar, 24BCSA37, CSE

- **Member 2:** Mrinalraj Arya, 24BCSA27, CSE

- **Member 3:** Sumit Kumar, 24BCSI13, CSE

- **Contact (any one member): 9097632962**

Data Strategy -

→ Did you use any external data beyond the provided train.csv?

   No. .

Data Cleaning and Preprocessing -

→ I replaced NULL with **mean** of that column

→ I used one hot encoding for columns with categorized value

→ And Nope I didn't removed any row because number of NULL values are more than 100 so removing rows can reduce the size of data.

→ Yes used one hot encoding and also scaled some columns for distance based models.

Feature Engineering

What were your top 3 most important features and why?

Rank	Feature	Why It Matters
1	DistanceKm	As Distance is Directly Proportional to Getting Delay
2	IsPeak Hour	In Co-Relation Matrix It's values was 0.077(Max)
3	RestaurantAvg	As Prep Time increases, the chance of delay increases.

Did you create any new features that were not in the original dataset?

Answer: No

Did you drop any features? Which ones and why?

Answer: I tried to drop some less related features like number of items but it didn't worked, So at the end I give up .

Did you check for correlations or feature interactions? What did you find?

Answer: Yes I checked and got to know which features are more important and which are less .

Model Selection and Training

What is your final model?

Answer: Random Forest Classifier

What other models did you try before this? Why did you pick the final one over them?

Model Tried	Validation Score	Why You Rejected / Kept It
KNN	61.22222	Quite Lower Accuracy than the currently used model
XGBOOST	59.87272	Lowest Accuracy

What hyperparameters did you tune? What values did you settle on?

Answer: An estimated 97 with a max depth of 17 and a random state of 42

How did you validate your model? (train/test split, cross-validation, etc.)

Answer: train → test → split

What was your best validation score before submitting?

Answer: 64.5

Honest Reflection

What did you try that did NOT work?

Answer: First Time I tried to mess the data because I directly replaced NULL with MEAN, that made different kind of object because NULL was included in MEAN . . . So I had to restart again.

What are the known limitations of your model?

Answer: Model Accuracy is not so good as it gave us a maximum accuracy of 64.5% .

What was the hardest part of this challenge for your team?

Answer: To increase the accuracy of the model while tuning the Model, we increased it from

52.3 → 64.5 but couldn't exceed that.

If you had 6 more hours, what would you do differently?

Answer: I will apply CHI-TEST to remove some column and I will try to increase the accuracy with better tuning .

Team Collaboration -

I Ankit Kumar, The team lead was well equipped with the "MachineLearning" Skills, did almost most of the parts of the project and my colleagues Mrinalraj Arya and Sumit Kumar did help me with the GitHub integrity and also could have helped me further to build the web-dashboard of the project when it is well developed to be scalable .

External Data Declaration -

Sorry But I used the data set that is duly provided internally .