261 FINAL PROJECT GANTT CHART

PROJECT TITLE	261 Final Project	COMPANY NAME	UC Berkeley School of Information
PROJECT MANAGER	Team 4-1	DATE	3/1/25

SPRING BREAK WEEK

No	TASK TITLE	DESCRIPTION	NOTES/ COMMENTS	TASK OWNER	START DATE	DUE DATE	DURATION	STATUS	PCT OF TASK COMPLETE
0	**Phase 0 - Finalize Teams: Due March 9th**		**PHASE LEADER**	Danielle Yoseloff		bcources Link			
1	**Phase I - Project Plan: Due March 16th**		**PHASE LEADER**	Mohamed Bakr		bcources Link			
2	**Phase II EDA, Baseline Pipeline on all data: Due April 6th**		**PHASE LEADER**	Shruti Gupta		bcources Link			
2.0	**Project Management**		**Note:** phase 2 - team leader responsibilities		3/17/25	4/6/25	20	Complete	100%
2.0.1	Assign/update Phase Leader for Phase II			Team	3/17/25	3/18/25	2	Complete	100%
2.0.2	Provide team and project meta information & Update Task Allocation.	Ensure that the Team ID and complete list of team members and project meta information (e.g., email) is provided and allocate responsibilities for each team member.		Shruti Gupta	3/17/25	3/18/25	2	Complete	100%
2.0.3	Summarize findings & create slides	Develop a compelling, well-structured presentation.		Team	3/26/25	4/2/25	7	Complete	100%
2.0.4	Integrate Work and Finalize Report	Collect inputs from all team members and structure them into a single notebook and presentation files.		Shruti Gupta	4/2/25	4/6/25	5	Complete	100%
2.0.5	Conduct Final Review	Review the completed notebook/ presentation and ensure proper submission.		Team	4/6/25	4/6/25	1	Complete	100%
2.0.6	Submit on bcources	Submit the final notebook/ presentation in bcources and add team members to the submission		Shruti Gupta	4/6/25	4/6/25	1	Complete	100%
2.1	**EDA**				4/6/25	4/6/25	1	Complete	100%
2.1.1	Assess dataset sizes & storage requirements	Examine the three-month, one-year, and full OTPW datasets.	Joins dataset task	Mohamed Bakr	3/17/25	3/18	2	Complete	100%
2.1.2	Investigate patterns & propose handling strategies	Identify and document missing data handling methods.	Erica/Shruti, Flights/Weather	Shruti Gupta	3/19/25	4/6/25	18	Complete	100%
2.1.3	Identify categorical & numerical features	Understand raw vs. engineered features & transformation needs.	Erica/Shruti, Flights/Weather	Shruti Gupta	3/19/25	4/6/25	18	Complete	100%
2.1.4	Create key visualizations	Generate histograms, scatter plots, and correlation heatmaps.	Erica/Shruti, Flights/Weather	Shruti Gupta	3/31/25	4/6/25	7	Complete	100%
2.1.5	Dimensionality Reduction Evaluation	Assess feature selection and dimensionality reduction impact (Apply LASSO/PCA if needed).	Feature Selection: Erica/Shruti, Flights/Weather	Shruti Gupta	3/17/25	4/6/25	20	On Hold	0%
2.2	**Feature Engineering**				3/31/25	4/6/25	7	Complete	100%
2.2.1	Engineer time-related predictors	Add recency, frequency, and other time-based features.		Danielle Yoseloff	3/31/25	4/6/25	7	Complete	100%
2.2.2	Apply feature transformations	Implement feature transformation techniques for categorical/text data (e.g., hashing trick, tf-idf, stopword removal, lemmatization, tokenization, etc..).		Shruti Gupta	3/17/25	4/6/25	20	On Hold	0%
2.3	**Baseline Pipeline Development**	*Friday 3/28 meeting to reassess division of work*	Needs to be done (not necessarily tuned) by Apr 2 in-class pres		3/25/25	4/2/25	8	Complete	100%
2.3.1	Choose baseline ML models	Justify the selection of logistic/linear regression and ensemble models.	Chosen based on Phase I feedback received 3/25	Erica Landreth	3/25/25	3/25/25	1	Complete	100%
2.3.2	Define cross-validation strategy	Ensure proper time-based data split (no data leakage).		Erica Landreth	3/30/25	4/4/25	5	Complete	100%
2.3.3	Develop ML pipeline with automation	Implement a structured, reproducible pipeline.		Erica Landreth	3/27/25	4/2/25	6	Complete	100%
2.3.4	Collect baseline model performance metrics	Measure accuracy, precision-recall, and other KPIs.		Erica Landreth	4/1/25	4/2/25	2	Complete	100%
2.4	**Hyperparameter Tuning & Optimization**	*Friday 3/28 meeting to reassess division of work*			4/4/25	4/6/25	3	Complete	100%
2.4.1	Optimize baseline models	Use GridSearchCV to fine-tune hyperparameters.		Shruti Gupta	4/4/25	4/6/25	3	Complete	100%
2.4.2	Compare hyperparameter impact	Analyze which tuning settings improve results.		Shruti Gupta	4/4/25	4/6/25	3	Complete	100%
2.5	**Time-Series Considerations**	*Friday 3/28 meeting to reassess division of work*			3/26/25	4/4/25	9	Complete	100%
2.5.1	Implement rolling-window cross-validation	Ensure validation follows time order.		Erica Landreth	3/30/25	4/4/25	5	Complete	100%
2.5.2	Examine seasonal patterns	Identify cyclic trends affecting model performance.	Look into Prophet!!	Erica Landreth	3/26/25	3/30/25	5	Complete	100%
2.6	**Extra Credit**				3/30/25	3/30/25	1	Complete	100%
2.6.1	Perform full joins of flight & weather data	Store in AWS storage, document process.		Mohamed Bakr	3/23/25	3/30/25	8	Complete	100%
-- Insert Rows above this line									
3	**Phase III - Advanced model architectures and loss functions: Due April 19th**		**PHASE LEADER**	Erica Landreth		bcources Link			
3.0	**Project Management**		**Note:** phase 3 - team leader responsibilities		4/7/25	4/20/25	14	In Progress	95%
3.0.1	Assign Phase Leader for Phase III	Document the leader responsible for Phase III coordination.		Team	4/7/25	4/7/25	1	Complete	100%
3.0.2	Provide team and project meta information & Update Task Allocation.	Ensure that the Team ID and complete list of team members and project meta information (e.g., email) is provided and allocate responsibilities for each team member.		Erica Landreth	4/7/25	4/19/25	13	Complete	100%
3.0.3	In-Class Presentation	Develop a structured, engaging presentation (refer to the rubric for details).		Team	4/12/25	4/16/25	5	Complete	100%
3.0.4	Final Project Abstract	Include everything covered in previous phases, plus the new experiments and the final model selected, as well as your final results (report the number!)		Shruti Gupta	4/19/25	4/20/25	2	Complete	100%
3.0.5	Integrate Work and Finalize Report	Collect inputs from all team members and structure them into a single notebook and presentation files.		Erica Landreth	4/14/25	4/20/25	7	Complete	100%
3.0.6	Conduct Final Review	Review the completed notebook/ presentation and ensure proper submission.		Team	4/16/25	4/20/25	5	Complete	100%
3.0.7	Submit on bcources	Submit the final notebook/ presentation in bcources and add team members to the submission		Erica Landreth	4/20/25	4/20/25	1	Not Started	0%
3.1	**Data and Feature Engineering**				4/7/25	4/19/25	13	Complete	100%
3.1.1	Summarize Data Lineage	Identify datasets used (flights, weather, etc.) and their origins. Describe key joins between datasets (e.g., flights with weather by timestamp).	MB: 5 year join	Mohamed Bakr	4/10/25	4/13/25	4	Complete	100%
3.1.2	List Feature Families	Group features into families (e.g., time-based, graph-based, location-based, airline-specific). Provide a rationale for each family and its potential impact on prediction.	Build off of Phase II report output	Erica Landreth	4/17/25	4/20/25	4	Complete	100%
3.1.3	List Features per Family	List each feature under its respective family. Provide a brief description of each feature's purpose.		Erica Landreth	4/20/25	4/19/25	0	Complete	100%
3.1.4	Exploratory Data Analysis (EDA)	Visualize feature distributions and assess data quality. Highlight insights from EDA, such as feature correlations or missing values.		Erica Landreth	4/14/25	4/19/25	6	Complete	100%

261 FINAL PROJECT GANTT CHART

3.1.5	Evaluate Feature Value	Run experiments to assess each feature's contribution to model performance Summarize results, showing performance change with and without each feature.		Danielle Yoseloff	4/16/25	4/19/25	4	Complete	100%
3.2	**Model Experimentation**				4/11/25	4/19/25	9	Complete	100%
3.2.1	Random Forest Implementation	Train, evaluate, and document performance across datasets.	Others helping with param tuning	Danielle Yoseloff	4/11/25	4/19/25	9	Complete	100%
3.2.2	GBDT Implementation	Apply boosting techniques, tune hyperparameters, and monitor performance.	Others helping with param tuning	Erica Landreth	4/11/25	4/19/25	9	Complete	100%
3.2.3	MLP Neural Network	Implement two architectures with early stopping and compare results.	Others helping with param tuning	Shruti Gupta	4/11/25	4/19/25	9	Complete	100%
3.2.4	Measure performance of all trained models	Evaluate on validation sets before blind test.	Hyperparam tuning notebook setup	Mohamed Bakr	4/14/25	4/19/25	6	Complete	100%
3.3	**Modeling Pipeline and Fine-Tuning**				4/7/25	4/19/25	13	Complete	100%
3.3.1	Visualize Modeling Pipeline	Draw a visual representation of the modeling pipeline, including preprocessing.		Danielle Yoseloff	4/19/25	4/19/25	1	Complete	100%
3.3.2	Record Hyperparameters	Document hyperparameters considered for each model. Note any special configurations (e.g., early stopping, learning rate).		Erica Landreth	4/17/25	4/19/25	3	Complete	100%
3.3.3	Specify Loss Function	Present the loss function with its components (e.g., data loss, regularization).	Already present in report draft	Danielle Yoseloff	4/7/25	4/7/25	1	Complete	100%
3.3.4	Track Experimentation	Record the number of experiments conducted. Keep a table of all experiments and their settings/results.		Erica Landreth	4/13/25	4/19/25	7	Complete	100%
3.3.5	Record Experiment Details	Document the initial model setup before optimizations. Track changes made in subsequent experiments. Highlight the final selected model and write a gap analysis of the final model.		Shruti Gupta	4/13/25	4/19/25	7	Complete	100%
3.3.6	Summarize Best Results	List the top 1–3 results with performance metrics.		Shruti Gupta	4/19/25	4/20/25	2	Complete	100%
3.3.7	Record Computational Setup	Specify computing resources used for each experiment. Track total runtime for each experiment.		Mohamed Bakr	4/10/25	4/19/25	10	Complete	100%
3.4	**Time-Based Cross-Validation**				4/11/25	4/19/25	9	Complete	100%
3.4.1	Implement cross-validation	Implement time-based CV using 2015–2018 data for training/validation. Reserve 2019 as a blind test set.		Erica Landreth	4/11/25	4/16/25	6	Complete	100%
3.4.2	Examine how trends affect model accuracy	Adjust pipeline accordingly.		Shruti Gupta	4/15/25	4/19/25	5	Complete	100%
3.5	**Pipeline Integrity & Leakage Prevention**				4/18/25	4/19/25	2	Complete	100%
3.5.1	Leakage Checks	Validate that only historical data is used during training.		Mohamed Bakr	4/18/25	4/19/25	2	Complete	100%
3.5.2	Pipeline Auditing	Implement tests to ensure no data leakage occurs.		Mohamed Bakr	4/18/25	4/19/25	2	Complete	100%
3.6	**Results and Conclusion**				4/16/25	4/20/25	5	Complete	100%
3.6.1	Discussion of results	Present an interpretation of key results. Do not discuss any outcomes not presented in the results part.		Shruti Gupta	4/18/25	4/20/25	3	Complete	100%
3.6.2	Restate Project Focus and Hypothesis	Summarize the project's aim and its importance. Clearly state the hypothesis being tested.		Shruti Gupta	4/19/25	4/20/25	2	Complete	100%
3.6.2	Highlight Best Features	Identify the most impactful features. State the top-performing model and its configuration. Summarize key hyperparameter choices and tuning insights.		Danielle Yoseloff	4/16/25	4/20/25	5	Complete	100%
3.6.2	Discuss Significance of Results	Analyze what the results mean in a practical context. Explain the implications of findings for stakeholders and real-world scenarios.		Shruti Gupta	4/19/25	4/20/25	2	Complete	100%
3.6.2	Discuss Future Directions.	Suggest improvements and further exploration. Recommend additional features, models, or data to enhance performance.		Shruti Gupta	4/16/25	4/20/25	5	Complete	100%
3.7	**Extra Credit**				4/7/25	4/19/25	13	On Hold	0%
3.7.1	Experiment with additional neural networks	Try advanced architectures for extra credit.		Danielle Yoseloff	4/7/25	4/19/25	13	On Hold	0%
3.7.2	Obtain and preprocess extra flight/weather data	Provide a clean dataset for Flights and/or Weather for Years 2020-2023		Mohamed Bakr	4/7/25	4/19/25	13	On Hold	0%
-- Insert Rows above this line