

Identifying Vendors using Topological data analysis...

| | |
|--|------|
| Preliminary research | 100% |
| Applications of Topological data analysis | 100% |
| A practical introduction to TDA | 100% |
| Research papers on Topology and geometrical analysis | 100% |
| Invoice and receipt application areas study | 100% |
| Documenting Literature Review | 100% |
| Data collection, preparation and EDA | 97% |
| Research datasets with bounding boxes (topology) | 100% |
| Kaggle invoice dataset with vendor names explore | 100% |
| YOLO data analysis methods and techniques | 100% |
| SROIE dataset collection | 100% |
| Exploratory data analysis on invoice SROIE data | 100% |
| Identify outliers and eliminate | 75% |
| Implementation (Main Bulk of Project) | 25% |
| ML approaches | 53% |
| Vendor identification using ML- Tutorial | 75% |
| Feature vector with area, centroid in sparse matrix | 80% |
| Prepare train and test datasets with topology for baseline | 80% |
| Baseline ML approach - Logistic regression | 75% |
| KNN based clustering | 45% |
| Hyperparameter tuning, class balance and preprocessing differently | 25% |
| Other ML algorithms like Random Forest | 20% |
| Achieve 80% or more accuracy | 0% |
| Deep Neural network based classifiers | 0% |
| Explore CNN or RNN | 0% |
| Implement data augmentation to balance vendor classes | 0% |
| Topological data analysis techniques | 0% |
| TDA methods and clustering- YouTube tutorial | 0% |
| Python TDA libraries explore | 0% |
| Implement Cornell University paper | 0% |
| Balance vendor classes and apply TDA algorithms | 0% |
| Eliminate Bias and Overfitting/ Underfitting | 0% |
| Results | 4% |
| Compare ML, DL and TDA approaches | 0% |
| Evaluate and Tabulate the results with metrics | 0% |
| MioU metric | 0% |
| Document results, discussion and conclusion | 20% |
| Data Innovation Showcase conference | 0% |
| Participate in NICD Innovation showcase | 0% |
| Represent Sage and produce a cost benefit analysis | 0% |
| Future work based on Vendor Fraud identification | 0% |
| Project Deliverables | 22% |
| Title and Description | 100% |
| Ethical approval | 100% |
| Interim report submission | 100% |
| Poster design | 0% |
| Poster submission | 0% |
| Video for Oral presentation | 0% |
| Final source code ready | 0% |
| Thesis report document and submission | 0% |

