|  |  |  |
| --- | --- | --- |
| Phase | Changes | Steps |
| 1. Scope reset | 3 ADE targets only | Columns to **keep** everywhere: smiles, label\_Gastrointestinal disorders, label\_Infections and infestations, label\_Nervous system disorders. Treat as multilabel; allow “none-of-three”. |
| 1. Dataset Curation | Lean 4 column dataframe | 1 ➜ Drop every other column found in the raw CSVs. 2 ➜ Delete rows with missing / empty SMILES. 3 ➜ Remove rows where all three labels are 0 **unless** you want extra negatives for contrastive pre-training. 4 ➜ Cast targets to uint8 and save single consolidated file data/interim/ade\_3lbl.csv. |
| 1. Split Strategy | Scaffold stratified train/val/test | *Across the* ***full*** *cleaned file*: • Generate Bemis–Murcko scaffold IDs (RDKit). • Stratify 70/15/15, preserving per-label prevalence. • Write the split column back to the CSV and persist scaffolds.pkl for audit. |
| 1. Tensor Build | From CSV to .npy pairs | For each split: python<br>tok = tokenizer(list(df.smiles), max\_length=128, padding='max\_length', truncation=True, return\_tensors='np')<br>X = tok['input\_ids']; y = df[targets].values<br>np.save('data/processed/X\_val.npy', X)<br>np.save('data/processed/y\_val.npy', y)<br> |
| 1. ChemBERTa fine tune (3-lable head ) | Balanced learning | |  | | --- | |  |  |  | | --- | | • Replace head → out\_dim = 3, sigmoid. • Loss: BCEWithLogitsLoss(pos\_weight=inverse\_freq).  • Metrics: per-label F1 **plus** “all-negative” accuracy. | |
| 1. Threshold tuning | Robust “none” decision | On the held-out **val** tensors, grid-search thresholds 0.05 – 0.5 for each label to maximise balanced accuracy. Save to thresholds.json. |
| 1. SHAP configuration | One cache per split | • Background: 100 diverse train SMILES embeddings. • Run shap.DeepExplainer once per split → shapes 3×B×128. • Save shap\_train/val/test.npz containing shap, y, token\_ids. |
| 1. SHAP post processing | Chemistry aware features | • Map tokens → functional groups (SMARTS). • Aggregate |
| 1. Meta explanation MLP | Structured rationale classes | • Input: 15-dim SHAP features (float32). • Output: multi-label vector over curated explanation classes (e.g. 15-20 common toxicophores). • Loss: BCEWithLogits; early stop on val macro-F1. Save model meta\_mlp.pt + scaler. |
| 1. UI & UX | Lean, cached workflow | • User SMILES → ChemBERTa → logits → thresholds. • Retrieve/compute SHAP → heat-map per label. • SHAP vector → meta-MLP → class IDs → sentence templates. • Show probabilities, colour-graded SMILES, and justification sentences; JSON export. |
| 1. Ops & monitoring | Reproducible & auditable | • Version raw CSVs + scripts with Git/DVC. • Store .npy and .npz artefacts under data/. • Log training metrics to W&B. • Monthly: regenerate splits + SHAP; retrain meta-MLP; watch drift (Tanimoto vs. training) and feedback on explanations. |