AGENDA – DAY 10 – 14-DEC-2025 (SUN)

- **Ensemble Learning**
- **CEPs**
- **Q & A**
- **SUMMARY & CLOSURE**

TECHNIQUE	ESSENCE	USE CASE	BENEFITS	CHALLENGES
HARD VOTING	MAJORITY OF BINARY VOTES	BEAST CANCER PRED (M OR BENIGN), LOAN APPROVAL DECISIONS	SIMPLE, TRANSPARENT	IGNORES CONFIDENCE LEVELS
SOFT VOTING	AEVRAGE PROBABILITY SCORES	CHURN PREDICTION	**BALANCED CONFIDENCE**	PORBABILITY OUTPUTS
WEIGHTED SOFT VORING	WEIGHTED AVERAGE OF PROBABILITIES	CREDIT ROSK SCORING WITH EXPERT WEIGHTING	PRIORITISES RELIABLE MODELS	CAREFUL WEIGHT TUNING

BANK: → FRAUD TRAN → DETECTION. $P(F) = 0.55$

→ 3 MODELS
→ LOG REG → 55% → FRAD
→ DT → 40% →
→ SVM → 70% →

HARD VOTING: LR → FRAUD ✓ (> 50%)

DT → NF (< 50%)

SVM → FRAUD ✓ (> 50%)

MAJORITY VOTING 2/3 → FRAUD

DIFFERENT CASE:

LR → 49% CHANCE OF FRAUD)
        ↳ "NF" → NF → NOT FRAUD

DT → 45%   "NF"

ENT|ML|(F1)
→LESS FALSE

SVM → 90%   "FRAUD"

HV → 2 SAY → "NF" → MAJORITY
                "NF"
        1 → "FRAUD"

$\frac{139}{95}$
$\frac{184}{3}$
61.3

SOFT VOTING: (49% + 45% + 90%)/3
        "FRAUD" = 61% ✓ → F
                            →NF

CEP 1 : Creating Cohorts of Songs:

- **INITIAL INSPECTION**
    - MISSING VALUES
    - DUPLICATES
    - OUTLIER DETECTION & TREATMENT …

- **DATA CLEANING :**
    - **STANDARDISE NUMERICAL FEATURES (Z-SCORE OR MIN-MAX) etc**

- **EDA**
    - **VISUALISATIONS**
        - POPULARITY DISTRIBUTION OF ALBUMS ACROSS THE YEARS
        - FEATURE CORRELATIONS – EXAMPLE, ENERGY & DANCEABILITY, ENERGY & LOUDNESS ETC
- DIMENSION REDUCTION

- CLUSTER ANALYSIS
    - PATTERN DISCOVERY
        - Identify the clusters – KMEANS, HIERARCHICAL, T-SNE ETC
        - EXPLORE FEATURE DISTRIBUTIONS THAT GUIDES YOUR CLUSTERING LOGIC
- DETERMINE OPTIMAL NUMBER OF CLUSTERS
    - ELBOW METHOS
    - SILHOUTTE SCORE …
- LABEL & INTERPRET THE CLUSTERS – BUSINESS RECOMMENDATION

NOTE: THE ABOVE SHOULD BE USED AS GUIDANCE ONLY