Tsetsgesuren Namjiljav

Opponent's review & evaluation of Tsetsegsuren Namjiljav (2026) — "Computational Approximation of Wine Expertise Using Object–Attribute Matrix Modelling and COCO-STD"

Based on the official Opponent's Evaluation Form (monographic thesis)

---

1) Use of AI & suspected plagiarism (risk assessment)

Documented AI use (positive: transparent)

- The thesis **explicitly documents** the use of an LLM as *support* (structure, wording, clarification) and includes a **separate benchmark session** with prompt + output in the annex.
- The author states boundaries: the LLM **was not used** to compute results, build COCO-STD models, or manipulate data; computations and implementation are claimed as independently executed.
- A responsibility statement claims the author retains full responsibility and that LLM use was supportive/comparative.

AI-authorship risk indicators (needs attention)

These are **not proof** of misconduct, but they are *signals* to check:

- **Highly repetitive formulaic phrasing** (e.g., repeated definitional expansions like "COCO-STD… that is, a COCO-based standard analysis…" across many sections) can be consistent with AI-assisted rewriting or templated writing.
- The annex contains a long LLM-generated "benchmark output" reproduced verbatim. This is acceptable **if clearly labeled** (it is), but it raises the importance of verifying that similar passages in the main text are original author writing.

Plagiarism suspicion (limited to what can be inferred from the PDF)

I cannot run an external plagiarism database check here. Based only on the document:

- The thesis includes many citations and a reference list, including the primary dataset source (Cortez et al., 2009) and multiple textbooks and web sources.
- However, several quotations appear **generic** and some references are **webpages / institutional pages**, which can make "patchwork paraphrase" harder to detect without a tool-based similarity scan.

Recommended verification actions (standard academic practice):

1. Ask for the **code repository / commit history** (Streamlit app + SQLite schema + run logs) and verify it matches described features (run_id logging, exports).
2. Ask for the **original Excel prototype** and at least one reproducibility run to confirm deterministic outputs and claimed metrics.
3. Conduct a short **oral technical check** (explain COCO-STD input, direction flags from correlation, and how MAE/correlation were computed).

Conclusion (AI & plagiarism):

AI use is **declared and bounded** (good academic transparency), but the writing style and the presence of large LLM text blocks warrant a **lightweight authorship verification** (recommended, not accusatory).

---

2) Scoring (per Opponent's form)

The form defines **max points** per category and the grade thresholds:

- Fail: 0–119
- Satisfactory: 120–139
- Average: 140–159
- Good: 160–179
- Excellent: 180–200

Proposed point allocation (with justification)

	Criterion	Max	Score	Rationale (very concise)
1	**Topic & Objectives**	40	**30**	Clear aim + objectives/tasks; relevance is plausible; some objectives over-ambitious vs evidence.
2	**Literature review**	40	**26**	Adequate quantity + "coverage matrix", but analysis is often definitional; heavy reliance on mixed-quality web/institutional sources; limited critical synthesis.
3	**Author's own work**	60	**44**	Strong implementation narrative (Excel → Streamlit), reproducibility/logging and scalability tests; but methodology validity and evaluation design are modest (subset-based, limited model comparison rigor).
4	**Structure, style, formal aspects**	40	**20**	Major formal issue: multiple "**Error! Bookmark not defined.**" in ToC; language is readable but repetitive; formatting inconsistencies.
5	**Overall impression**	20	**12**	Solid applied project with transparent pipeline; academic polish and validation depth need strengthening.

Total: 132 / 200 → Satisfactory

3) Detailed evaluation by criterion

A) Topic & Objectives (max 40) — 30/40

Strengths

- The thesis states a **clear research aim**: computational approximation of wine expertise using OAM + COCO-STD, and it enumerates explicit objectives linked to chapters.
- The project framing is pragmatic: decision-support rather than replacing expert tasters.

Weaknesses / gaps

- Some objectives imply stronger empirical claims than the evaluation supports (e.g., "approximate expertise" could be interpreted as predictive modelling, but validation is relatively modest).
- The "informational added-value" scenario uses assumed parameters (e.g., "40% workload reduction") without empirical justification.

Improvement suggestions

- Tighten objective wording to distinguish **concept-testing** vs **predictive performance** and explicitly define what "approximation" means (e.g., acceptable MAE / correlation thresholds).

B) Review of the Literature (max 40) — 26/40

Strengths

- The thesis includes a structured literature chapter and explicitly connects foundational concepts (wine quality modelling, similarity analysis, OAM, COCO-STD, evaluation metrics).
- The author creates a **reference coverage matrix (T01–T16)** to demonstrate breadth across language/recency/source type.

Weaknesses

- The review frequently reads like **a chain of definitions and quotations** rather than a critical synthesis (e.g., many textbook "what is X" passages).
- The "coverage matrix" approach risks becoming a **box-ticking exercise**: one source per cell is not evidence of deep coverage; quality varies substantially across sources (textbooks vs generic web pages).
- Limited engagement with mainstream methods on the same dataset (e.g., baseline regression/trees, cross-validation practices) beyond general mentions; the thesis would benefit from positioning COCO-STD relative to standard predictive approaches more concretely.

Improvement suggestions

- Replace part of the definitional material with:
  a. **what prior work achieved on this dataset**,
  b. **why COCO-STD/OAM adds value** (transparency, auditability), and
  c. **limitations** vs conventional ML.

C) Presentation of the Author's Own Work (max 60) — 44/60

Strengths (substantial)

- Clear "own developments" focus: **OAM construction**, **rank transformation**, **COCO-STD processing**, KPI computation, deviation interpretation, and a **Streamlit application** with a Help/manual and exports.
- Strong reproducibility intent: run-level traceability using **UUID run_id**, **SQLite logging**, and per-run export folders.
- Scalability evidence includes repeated runs and summary runtime stats for several N values.

Weaknesses (method + evaluation rigor)

- Validation depth is limited: the headline relationship is **Pearson r ≈ 0.4413** (moderate) on a **100-sample subset**, and the thesis reports MAE (0.554) but does not demonstrate robust out-of-sample evaluation design in the main narrative.
- The work emphasizes deterministic pipelines, but the academic question ("approximation of expertise") would benefit from at least:
  - a simple baseline (mean/median predictor) and
  - a train/test or cross-validation framing, to show incremental value.
- The "MORE valued / LESS valued" rule is deterministic and useful for decision support, but interpretation should be more cautious: it may reflect model structure rather than true expert error. The thesis does note this, but it remains a key risk in practical interpretation.

Improvement suggestions

- Add a minimal, thesis-appropriate evaluation package: baseline comparisons + cross-validation, and report "within ±1" accuracy (common for ordinal ratings). This would strengthen the empirical claim while staying consistent with transparency goals.

D) Structure, Style, and Formal Aspects (max 40) — 20/40

Major formal problem

- The Table of Contents shows multiple **"Error! Bookmark not defined."** entries for Chapter 4–5 subsections. This is a serious formatting defect for a submitted thesis.

Additional issues

- Style is generally clear, but there is **noticeable repetition** and "template-like" phrasing.
- Minor inconsistencies: some lists/titles formatting, occasional awkward phrasing, and overly long definitional sentences.

Required fixes (before archival)

- Regenerate ToC properly (Word bookmarks / cross-references), ensure all headings resolve.
- Light copy-edit pass: reduce repetitive definitional expansions; tighten language and remove redundancy.

E) Overall Impression (max 20) — 12/20

- The thesis is **strongest** as an applied, reproducible decision-support prototype (Excel → Streamlit + SQLite logging) with explicit pipeline stages and usability considerations.
- It is **weaker** as an academic empirical study because evaluation design and literature synthesis are less rigorous than the engineering narrative.
- With formal corrections and a slightly stronger validation section (even minimal baselines + CV), this could move toward the "Average/Good" range.

Final result

✅ **Total: 132 / 200 points → "Satisfactory"** (per form thresholds)

If you want, I can also output this in the exact Opponent's form style (short bullet comments per section + one-paragraph summary + final grade line).