UEM24 at NTCIR-18 MedNLP-CHAT

A Machine Learning Approach to Multilingual Healthcare Risk Prediction

Ayantika Das[1], Anupam Mondal[2]

[1]Brainware University, University of Engineering & Management, West Bengal, India
ayd.cse@brainwareuniversity.ac.in

[2]Institute of Engineering and Management, University of Engineering & Management, West Bengal, India
anupam.mondal@iem.edu.in

Abstract

Risk prediction in healthcare conversations is vital for safety and compliance. We developed machine learning models to detect medical, legal, and ethical risks from multilingual patient-doctor interactions. Using translated datasets, we applied text preprocessing and trained models for classification and regression to support safer healthcare communication.

Motivation

1.Limited and imbalanced multilingual healthcare data challenges accurate risk detection.

2.Lack of subjective labels in Task2 complicates reliable quality assessment.

3.Complex medical language and potential annotation biases demand robust preprocessing and modeling.

Objective

1.Classify **objective risks** (TRUE/FALSE) and predict **subjective scores** (fluency, helpfulness, harmlessness).

2.Address multilingual, domain-specific challenges with robust models for limited, imbalanced data.

3.Provide reliable risk assessment to enhance patient safety and compliance.

Dataset

1.Combined Task1 & Task2 (100 samples each), total 200 QA pairs with objective risk labels: **medicalRisk**, **ethicalRisk**, **legalRisk**.

2.Task1 has subjective scores (**fluency**, **helpfulness**, **harmlessness**) from 79 annotators on a -2 to +2 scale, stored as distributions.

3.Task2 lacks subjective scores; predicted using regressors trained on Task1 (Gradient Boosting, XGBoost).

4.Final dataset merges both with objective and subjective labels.

Preprocessing

1.Lowercasing, special characters and punctuation removal

2.Tokenization, stopword removal, lemmatization

3.CountVectorizer (1–3 grams, 5000 features)

Models

Task1:

1.**Medical & Ethical Risks:** Logistic Regression—simple and effective for binary classification.

2.**Legal Risk:** Nu-SVC (nu=0.1)—better handling of class imbalance.

3.**Subjective Scores:**

i)**Fluency & Helpfulness:** Gradient Boosting

ii)**Harmlessness:** XGBoost

Task2:

1.**Medical Risk:** Logistic Regression

2.**Legal & Ethical Risks:** Nu-SVC (nu=0.1)

3.Subjective scores predicted using Task1 regressors

System Architecture

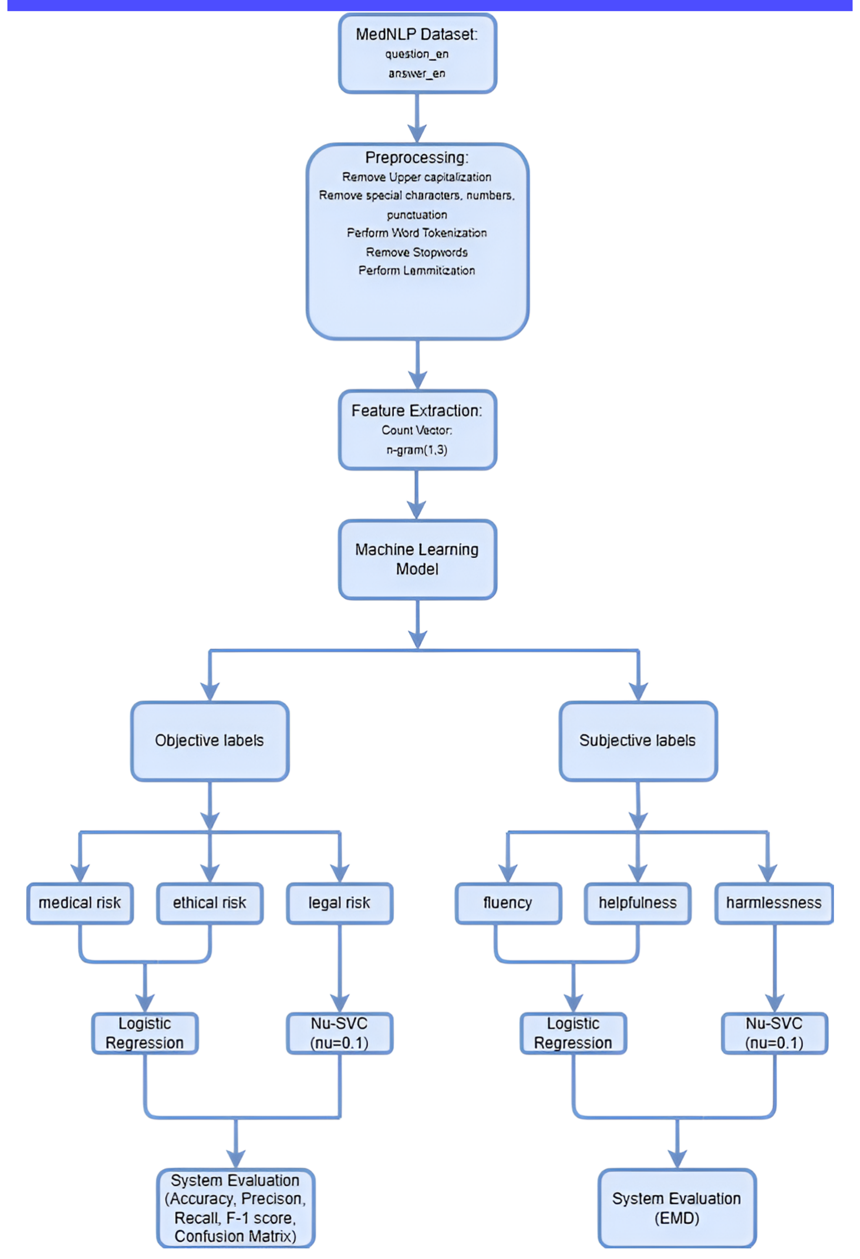

Figure 1: Pipeline of the proposed system

Task1 Results

Objective Labels:

Label	Acc	Prec	Rec	F1
Medical	0.595	0.557	0.531	0.50
Legal	0.786	0.423	0.458	0.44
Ethical	0.897	0.585	0.595	0.59

Subjective Scores (EMD):

- Fluency: 0.012
- Helpfulness: 0.018
- Harmlessness: 0.016

Task2 Results

Label	Acc	Prec	Rec	F1
Medical	0.634	0.612	0.596	0.594
Legal	0.750	0.700	0.646	0.658
Ethical	0.625	0.620	0.625	0.619

Discussion & Conclusion

1.Logistic Regression and Nu-SVC performed well for different risks.

2.Subjective scores predicted accurately, especially fluency.

3.Some overfitting in Task1; overall strong ML results.

4.Future work: deep learning, better balance, and more features.

References

[1] E. Aramaki et al., 2016. Overview of the NTCIR-12 MedNLP Doc Task. NTCIR-12 Conf.

[2] E. Aramaki et al., 2025. NTCIR-18 MedNLP CHAT Task Overview. NTCIR-18 Conf.

[3] V. Chaturvedi et al., 2020. A Supervised Approach to Analyse and Simplify Micro-texts. IEM Graph 2018, Springer.

[4] A. Mondal et al., 2018. Relation Extraction of Medical Concepts. Cognitive Computation.

[5] A. Mondal et al., 2015. Lexical Resource for Medical Events. IEEE ICDMW 2015.

[6] A. Mondal and D. Das, 2021. Ensemble Approach for Medical Concepts. Sādhanā.

[7] S. Wakamiya et al., 2017. Overview of the NTCIR-13 MedWeb Task. NTCIR-13 Conf.

[8] S. Wakamiya et al., 2023. NTCIR-17 mednlp-sc Adverse Drug Event Detection. NTCIR Conf.

[9] S. Yada et al., 2022. Real MedNLP: Document-Based NLP Task. NTCIR-16 Conf.