

An Explainable AI Framework for Personalized Astrology Consultation Recommendation Using Machine Learning and Natural Language Processing

Saloni Sandeep Shelar¹ and Ms. Yogita Kadbane²

¹Student, Master of Computer Applications, Sadhu

²Assistant Professor, Master of Computer Applications,
Sadhu Vaswani Institute of Management Studies for Girls, Pune,
Savitribai Phule Pune University, Maharashtra, India

Abstract: Astrology consultations usually begin with understanding a person's birth details and identifying the area of life for which guidance is required. When many consultation requests are received, this initial categorization can become time-consuming. This study presents an Explainable Artificial Intelligence (XAI) framework that uses Machine Learning (ML) and Natural Language Processing (NLP) to recommend an appropriate astrology consultation category.

The frame work combines the user's natural-language consultation query with structured information derived from birth details. The structured representation includes Lagna, planetary positions, Nakshatra, Nakshatra Pada, Dasha information, age, and gender. TF-IDF is used to represent consultation queries, and the resulting text features are combined with categorical and numerical astrological features.

Decision Tree, Random Forest, and XGBoost are compared. XGBoost gives the strongest result on the current held-out test set, with 95.00% accuracy, 96.25% weighted precision, 95.00% weighted recall, and 95.00% weighted F1-score. SHAP and LIME are used to explain individual predictions. For the example query "Should I change my current job?", the model predicts Career with 92.88% confidence.

A strict unseen-template evaluation gives 58.27% accuracy, indicating that substantially different wording remains challenging. The study therefore presents the framework as a consultation-category recommendation and decision-support approach rather than a scientific proof of astrological relationships. Independent validation using genuine anonymized customer consultation records remains important.

Keywords: Astrology Consultation; Machine Learning; Natural Language Processing; Explainable AI; XGBoost; SHAP; LIME; Personalized Recommendation

I. INTRODUCTION

People seek astrology consultations for different areas of life, including career, marriage, education, finance, health, and business. Before a consultation is conducted, the astrologer generally needs to understand the question and review relevant birth-chart information. In a manual process, the initial categorization depends on the experience of the person handling the request.

Artificial Intelligence can assist with this initial step. NLP can analyse how a user expresses a concern, while supervised Machine Learning can learn patterns from previous consultation records and recommend a consultation



category. The proposed framework also uses structured astrological variables so that the user's question is considered together with birth-related context.

The purpose is not to replace an astrologer or establish scientific causality. Instead, the system is designed as a decision-support layer that can help organise consultation requests and provide an interpretable reason for the recommended category.

1.1 Objectives of the Study

The study aims to develop and evaluate a Machine Learning framework for consultation-category recommendation using natural-language queries and structured astrological information. The work specifically addresses NLP-based query processing, generation of astrological features, hybrid feature integration, comparison of Decision Tree, Random Forest and XGBoost, evaluation using standard classification metrics, SHAP and LIME explanations, strict unseen-template generalization testing, and independent primary-data validation.

II. LITERATURE REVIEW

Research on machine learning classification shows that tree-based methods remain useful for structured and mixed feature spaces. Random Forest takes an average of multiple choice trees to enhance robustness and XGBoost employs gradient boosting to create a strong sequence of trees (Breiman, 2001; Chen & Guestrin, 2016). Decision Tree Learning is a good baseline for classification problems where the splits are interpretable and based on features (Quinlan, 1986).

TF-IDF is also a valid representation for text classification, as it transforms a document or query into weighted numerical features based on term relevance (Ramos, 2003; Salton & Buckley, 1988). Although BM25 is not used as the main representation in the current study, retrieval based work also brings up the importance of term weighting and document/query matching (Robertson & Zaragoza, 2009).

The need to interpret predictions has led to techniques such as SHAP (Lundberg & Lee, 2017) that assigns feature-attribution values, and LIME (Ribeiro et al., 2016) that builds a local interpretable approximation around an individual prediction. Other work has discussed the evaluation of interpretability itself, and the difficulty of providing rigorous explanations of complex models (Baehrens et al., 2010; Doshi-Velez & Kim, 2017; Sundarara.

Modern NLP has progressed from distributed word representations to contextual transformer models such as ELMo, BERT, RoBERTa, DistilBERT, and the original Transformer architecture (Peters et al., 2018; Devlin et al., 2019; Liu et al., 2019; Sanh et al., 2019; Vaswani et al., 2017). The present work intentionally uses TF-IDF as a transparent baseline rather than a large pretrained language model. Deep-learning research has nevertheless established that representation learning can capture complex patterns when sufficiently large data are available (LeCun et al., 2015; Mikolov et al., 2013; Pennington et al., 2014).

Recommender-system research also supports the use of hybrid representations and contextual signals when a user expresses a need in free-form language (Lu et al., 2015; He et al., 2017). These ideas are relevant to consultation recommendation because the query directly expresses the user's area of concern.

III. RESEARCH GAPS

Previous astrology related research has explored profession prediction, event prediction and other horoscope based classification tasks. Recent works include machine learning approaches for astrological prediction (Vigneswaran & Nagadeepa, 2022) and deep learning based profession prediction (Kanade & Kanade, 2023; Jaiganesh & Parameswari, 2024a, 2024b; Jajoo et al., 2026). However, there is limited work that brings together a user's natural-language consultation query, structured astrological information, consultation-category recommendation, and model explanations within one framework.

The present study addresses this gap by integrating NLP, structured birth-chart features, supervised classification, SHAP and LIME, while explicitly testing performance on unseen query wording.



IV. RESEARCH METHODOLOGY

The research workflow follows these stages: data collection, preprocessing, NLP processing, astrological feature generation, feature integration, model training, model evaluation, XAI interpretation, generalization testing, and primary-data validation.

4.1 Data Collection

The initial model-development experiment used the publicly available Jyotish-LLM astrology dataset from Hugging Face as secondary data. The study task was defined over six consultation categories: Career, Marriage, Health, Finance, Education, and Business. The secondary dataset was used to develop and evaluate the initial model. A separate set of genuine anonymized customer consultation responses was reserved for primary validation and was not used to inflate the development result.

A separate 100-row synthetic test workbook was used only for pipeline testing and demonstration. It is not treated as 100 genuine customers and is not presented as primary research evidence.

4.2 Data Preprocessing

The data-processing stage included checking missing values and duplicates, cleaning consultation queries, encoding categorical variables, scaling numerical variables, excluding personally identifiable information from model inputs, and separating the consultation category as the target variable.

4.3 NLP Processing

Consultation questions were converted into TF-IDF word n-gram features. The final NLP representation contained 251 features. The representation captures terms and short word combinations that frequently distinguish categories, such as employment and career terminology in a career-related question.

4.4 Astrological Feature Generation

Birth date, time, and place were processed using a deterministic astrological feature-generation component. The framework used Lahiri sidereal calculations and derived Lagna, planetary Rashi, Nakshatra, Nakshatra Pada, planetary longitudes, Dasha information, age, and gender. This feature-generation layer is separate from the ML classifier.

4.5 Feature Integration and Model Development

The final hybrid representation contained 251 NLP features, 446 encoded categorical features, and 12 numerical features, giving a total of 709 features. Decision Tree, Random Forest, and XGBoost were trained within the same experimental framework. The final selected XGBoost configuration used 150 estimators, maximum depth 2, learning rate 0.05, minimum child weight 1, subsample 1.0, and column subsampling 0.8, with random state 42.

4.6 Experimental Design and Evaluation

The primary model-development dataset was divided using a stratified 80:20 split with random state 42. The held-out test set contained 20 records and was kept untouched for final evaluation. Accuracy, weighted precision, weighted recall, and weighted F1-score were used as the main metrics. An additional strict evaluation was performed using previously unseen consultation-query templates to assess generalization.



V. RESULTS AND DISCUSSION

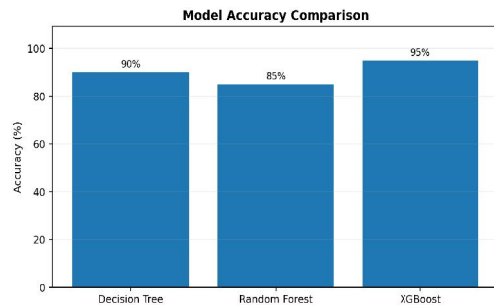


Figure 2. Accuracy Comparison of Machine Learning Models

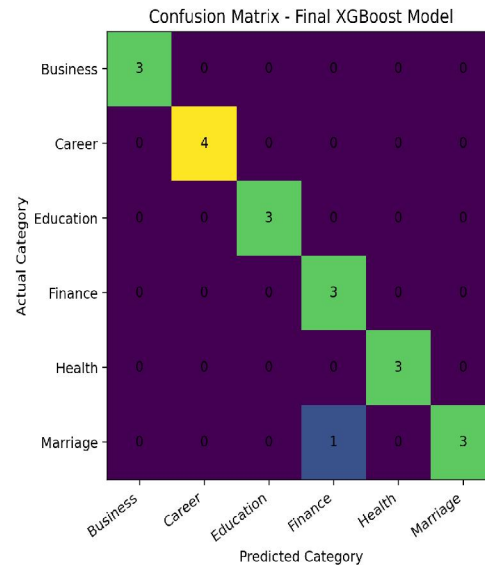


Figure 3. Confusion Matrix of the Final XGBoost Model

The three models produced different classification performance. Decision Tree achieved 90.00% accuracy, Random Forest achieved 85.00%, and XGBoost achieved 95.00%. XGBoost was therefore selected as the final model for the current experiment.

The confusion matrix has 20 held out test records. All of the Business, Career, Education, Finance and Health were correctly classified in the test set. For Marriage, 75% of the records were correctly classified and one was classified as Finance. So the final model predicted 19 out of 20 correctly.

The final XGBoost result was 95.00% accuracy, 96.25% weighted precision, 95.00% weighted recall, and 95.00% weighted F1-score. Because the test set is small, these figures should be treated as an experimental result rather than a definitive real-world accuracy estimate.

5.1 Category-wise Performance and Feature Groups

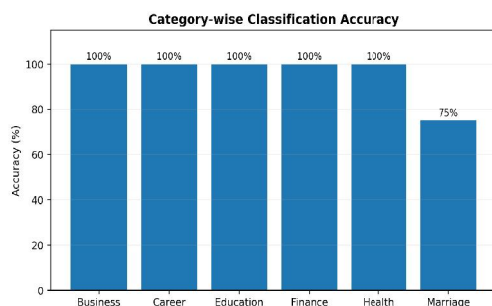


Figure 4. Category-wise Classification Accuracy

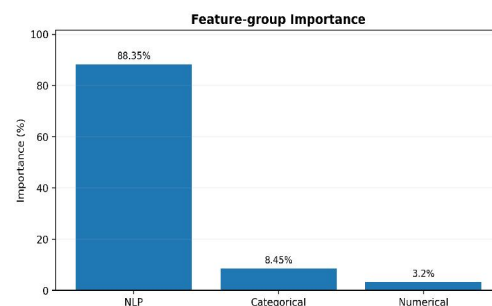


Figure 5. Relative Importance of Feature Groups



Category-wise accuracy was 100% for Business, Career, Education, Finance, and Health, while Marriage reached 75% because of the single Marriage-to-Finance error in the test set.

Feature-group importance was dominated by the NLP representation (88.35%), followed by categorical features (8.45%) and numerical features (3.20%). These percentages describe model feature importance and should not be interpreted as contributions to accuracy.

VI. EXPLAINABLE AI INTERPRETATION

The final XGBoost model was analysed with SHAP and LIME to understand how individual features influenced predictions. The purpose of the explanations is to make the classification process more transparent and to show the evidence used by the model for an individual case.

6.1 SHAP Explanation

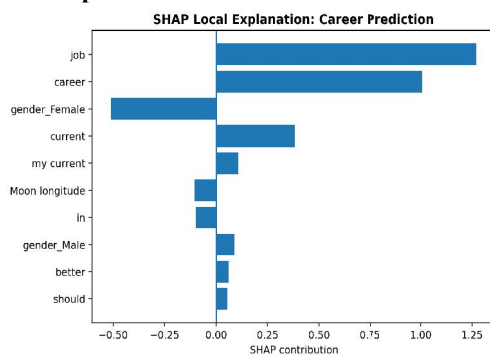


Figure 6. SHAP-based Local Explanation

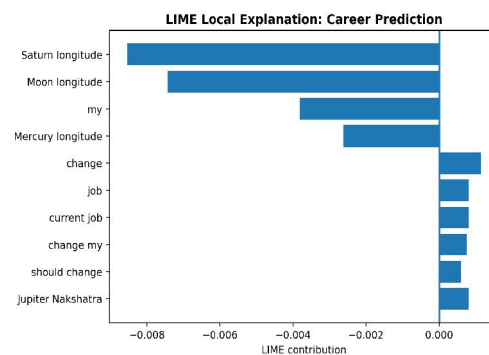


Figure 7. LIME-based Local Explanation

For the consultation query “Should I change my current job?”, the final model predicted Career with 92.88% confidence. The SHAP explanation showed strong positive contributions from terms associated with job, career, current, and my current. Some structured features pushed in the opposite direction but had smaller effects.

LIME provided a complementary local explanation for the same prediction. Textual features related to change, job, and current job supported the Career class, while selected numerical and categorical features made smaller contributions.

The two methods identified five common important features in the selected top-ranked sets. The top-20 overlap was 27.78% for SHAP and 25.00% for LIME, with 60.00% direction agreement. The difference illustrates that explanation methods can highlight different aspects of the same local decision.



VII. GENERALIZATION ANALYSIS

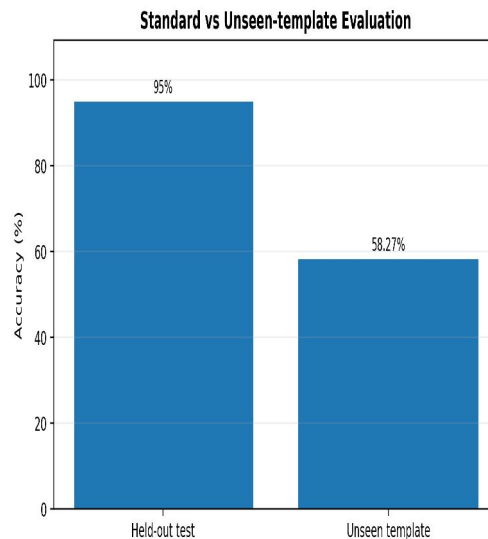


Figure 8. Model Performance under Different Evaluation Settings

A standard held-out test can overestimate practical performance when new questions resemble patterns seen during model development. To examine this issue, a strict unseen-template evaluation was performed.

The standard held-out accuracy was 95.00%, whereas the strict unseen-template accuracy was 58.27%. The drop indicates that the model benefits from wording patterns learned from the development data and has more difficulty when the same consultation intent is expressed in substantially different language.

VIII. PRIMARY DATA VALIDATION

The primary-data stage is intentionally separate from model development. The public Hugging Face dataset is secondary data, while genuine customer consultation responses collected for the study constitute primary data. The frozen final XGBoost model can be applied to these independent records without retraining.

The current primary-validation dataset contains a sample of genuine anonymized customer consultation records. Primary validation is intended to report accuracy, precision, recall, F1-score, a confusion matrix, and category-wise performance. No primary metric is inserted here unless it has been computed directly from the genuine records.

IX. LIMITATIONS

The main limitations are the small 20-record held-out test set, the use of secondary/model-development data for the initial experiment, the 58.27% strict unseen-template result, and the need for a larger and more diverse genuine customer dataset. SHAP and LIME explain the behaviour of the ML model; they do not establish scientific causality for astrological relationships.

Privacy is also important because genuine consultation records can contain personal information. Personally identifiable information should be removed before modelling, and the system should be presented as consultation-category decision support rather than a replacement for an astrologer or a medical diagnosis system.

X. CONCLUSION AND FUTURE WORK

This study presented an Explainable AI framework for personalized astrology consultation-category recommendation using NLP and Machine Learning. The framework combines free-form consultation questions with structured astrological features and compares Decision Tree, Random Forest and XGBoost models.



The current best held-out performance was obtained by XGBoost achieving 95.00% accuracy, 96.25% weighted precision, 95.00% weighted recall and 95.00% weighted F1-score. The confusion matrix shows 19 correct predictions out of 20, with the only error being a Marriage query classified as Finance. SHAP and LIME provided local explanations for a Career example, while the strict unseen-template evaluation showed that generalization remains a major challenge.

The next stage is to validate the frozen model on a larger sample of genuine customer records, increase the diversity of consultation questions, improve performance on unseen wording, compare TF-IDF with transformer-based NLP, examine additional ensemble models, and evaluate the system with practicing astrologers. A standalone user interface and later platform integration can follow after primary validation.

The findings therefore support the framework as an interpretable research prototype for consultation-category recommendation, while emphasizing that the reported 95.00% accuracy should not be treated as final real-world performance.

REFERENCES

1. Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 785–794). ACM. <https://doi.org/10.1145/2939672.2939785>
2. Breiman, L. (2001). Random forests. *Machine Learning*, 45, 5–32. <https://doi.org/10.1023/A:1010933404324>
3. Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning*, 1, 81–106. <https://doi.org/10.1007/BF00116251>
4. Ramos, J. (2003). Using TF-IDF to determine word relevance in document queries. In Proceedings of the First Instructional Conference on Machine Learning.
5. Salton, G., & Buckley, C. (1988). Term-weighting approaches in automatic text retrieval. *Information Processing & Management*, 24(5), 513–523.
6. Robertson, S., & Zaragoza, H. (2009). The probabilistic relevance framework: BM25 and beyond. *Foundations and Trends in Information Retrieval*, 3(4), 333–389.
7. Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems*, 30.
8. Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). Why should I trust you? Explaining the predictions of any classifier. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 1135–1144). ACM. <https://doi.org/10.1145/2939672.2939778>
9. Pedregosa, F., Varoquaux, G., Gramfort, A., et al. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.
10. Wolf, T., Debut, L., Sanh, V., et al. (2020). Transformers: State-of-the-art natural language processing. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations (pp. 38–45). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2020.emnlp-demos.6>
11. Vaswani, A., Shazeer, N., Parmar, N., et al. (2017). Attention is all you need. In *Advances in Neural Information Processing Systems*, 30.
12. Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In Proceedings of NAACL-HLT (pp. 4171–4186). Association for Computational Linguistics. <https://doi.org/10.18653/v1/N19-1423>
13. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521, 436–444. <https://doi.org/10.1038/nature14539>
14. Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. In *International Conference on Learning Representations*.



15. Pennington, J., Socher, R., & Manning, C. D. (2014). GloVe: Global vectors for word representation. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (pp. 1532–1543). Association for Computational Linguistics.
16. Peters, M. E., Neumann, M., Iyyer, M., et al. (2018). Deep contextualized word representations. In Proceedings of NAACL-HLT (pp. 2227–2237). Association for Computational Linguistics.
17. Liu, Y., Ott, M., Goyal, N., et al. (2019). RoBERTa: A robustly optimized BERT pretraining approach. arXiv:1907.11692.
18. Sanh, V., Debut, L., Chaumond, J., & Wolf, T. (2019). DistilBERT, a distilled version of BERT: Smaller, faster, cheaper and lighter. arXiv:1910.01108.
19. Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. arXiv:1702.08608.
20. Sundararajan, M., Taly, A., & Yan, Q. (2017). Axiomatic attribution for deep networks. In Proceedings of the 34th International Conference on Machine Learning (pp. 3319–3328). PMLR.
21. Baehrens, D., Schroeter, T., Harmeling, S., et al. (2010). How to explain individual classification decisions. Journal of Machine Learning Research, 11, 1803–1831.
22. Lu, J., Wu, D., Mao, M., Wang, W., & Zhang, G. (2015). Recommender system application developments: A survey. Decision Support Systems, 74, 12–32.
23. He, X., Liao, L., Zhang, H., Nie, L., Hu, X., & Chua, T.-S. (2017). Neural collaborative filtering. In Proceedings of the 26th International Conference on World Wide Web (pp. 173–182). International World Wide Web Conferences Steering Committee.
24. Vigneswaran, L. and Nagadeepa, N., 2022. Prediction of research studies status on astrological data using machine learning (ML) classification techniques. International Journal of Health Sciences, 6(S6), 9696–9710. <https://doi.org/10.53730/ijhs.v6nS6.12531>
25. Kanade, A., & Kanade, S. (2023). Astrological predictions using artificial intelligence and machine learning. Journal of Harbin Engineering University, 44(7).
26. Jaiganesh, S., & Parameswari, P. (2024a). An improvised learning model with multi-layered CNN for customized astrological prediction system. International Journal of Intelligent Systems and Applications in Engineering, 12(21s), 2489–2499.
27. Jaiganesh, S., & Parameswari, P. (2024b). An automated profession prediction application using the horoscope with deep learning-based strategies. In 2024 International Conference on Integration of Emerging Technologies for the Digital World (ICIETDW). IEEE. <https://doi.org/10.1109/ICIETDW61607.2024.10941579>
28. Jajoo, V., Verma, D., & Gupta, G. (2026). Astrological prediction of entrepreneurial potential using machine learning classification techniques. Siddhant: A Journal of Decision Making, 25(4), 255–259. <https://doi.org/10.5958/2231-0657.2025.00048.5>

