

A Hybrid Probabilistic Model for Latent Momentum and Match Outcome

Fashad Ahmed Siddique

Master's Degree in Artificial Intelligence, University of Bologna
fashad.ahmedsiddique@studio.unibo.it

November 22, 2025

Abstract

We present a hybrid probabilistic model combining Hidden Markov Models (HMMs) for momentum inference with Bayesian Networks (BNs) for match outcome prediction. We compared three BN structures: Naive Bayes (baseline), Expert DAG (knowledge driven), and Learned DAG via HillClimb (data driven). Results show Naive Bayes provided the most robust predictions (49.07% accuracy, 1.0139 log loss), tying with the Expert model. The complex Learned structure performed slightly worse (48.60%), suggesting that complex structure learning can lead to overfitting in stochastic domains like football.

Introduction

Domain

Football prediction involves inherent uncertainty due to the stochastic and highly variable nature of match outcomes. This complexity arises from numerous interacting factors, including team form, player availability, and random in game events that can influence results. Bayesian Networks provide a principled framework for modeling these uncertainties by encoding conditional dependencies between variables.

Aim

This work compares knowledge driven (Expert DAG) versus data driven (Learned DAG via HillClimb) structure learning, evaluating whether domain expertise or automated learning produces better predictive models. We also assess the value of latent momentum features inferred via HMMs in improving prediction accuracy compared to static features alone.

Method

The dataset used in this study is derived from the European Soccer Database (2000–2025) and consists of more than 230,000 matches. The modelling pipeline consists of the following stages: (1) Train 5 state Categorical HMM on sequences of match results (W/D/L) to infer momentum. (2) Discretize Elo ratings and momentum states into 5 bins. (3) Compare three BN structures:

Naive Bayes (conditional independence), Expert DAG (causal links), and Learned DAG (HillClimb with BIC). All models use Maximum Likelihood Estimation and Variable Elimination for inference, evaluated on time based train/test split (80/20).

Results

Naive Bayes achieved best Log Loss (1.0139) with 49.07% accuracy, tying with Expert and outperforming Learned (48.60%). All models struggle to predict draws, consistently favoring home or away wins (the “Draw Blindspot”), likely because draws rarely maximize the posterior probability in a 3 class distribution even when their likelihood is elevated. Comparison of exact inference (Variable Elimination) with approximate inference (Gibbs Sampling) showed convergence at N=500 samples. This confirms that for this dataset, simpler structures generalize better than complex data driven ones.

Model



Figure 1: BN Structures: Naive (Left), Expert (Center), Learned (Right).

Figure 1 illustrates the three BN architectures. Nodes represent: HomeElo, AwayElo, HomeMomentum, AwayMomentum (features) and FTR (Full Time Result, target). The Naive Bayes structure assumes all features are conditionally independent given FTR, providing a low variance baseline. The Expert DAG encodes domain knowledge with causal links (Elo influences Momentum, both influence FTR). The Learned DAG (via HillClimb) discovered complex dependencies from data by maximizing BIC score, but lacked intuitive causal structure and showed signs of overfitting.

Analysis

Experimental setup

The dataset used in this study comprises over 230,000 matches from the European Soccer Database covering the period 2000–2025. The feature set includes Elo ratings, computed from historical match outcomes, and momentum states inferred using a five state Categorical Hidden Markov Model (HMM) for each team. All continuous variables were discretized using quantile based binning into five categories to ensure balanced class distributions. The data were partitioned into a temporally ordered 80/20 train-test split to prevent data leakage and preserve the chronological structure of match sequences. Model performance was evaluated using accuracy, log loss, and confusion matrices. To better understand the contribution of the momentum features, we examined the latent states produced by the five state Categorical HMMs. Analysis of the state conditioned distributions of match outcomes and goal differences revealed that several states corresponded to interpretable performance regimes, such as sustained high form or low form periods. Examination of the learned transition matrices further showed strong self transition tendencies, indicating that teams tend to remain in a given momentum state across multiple matches. These findings support the interpretability and stability of the momentum features incorporated into the Bayesian Network models.

Results

The Naive Bayes model achieved the best log loss (1.0139) and an accuracy of 49.07%, tying with Expert and outperforming Learned (48.60%). All models struggle to predict draws, consistently favoring home or away wins (the “Draw Blindspot”), likely because draws rarely maximize the posterior probability in a 3 class distribution even when their likelihood is elevated. Comparison of exact inference (Variable Elimination) with approximate inference (Gibbs Sampling) showed convergence at N=500 samples. To contextualise these results, two baseline predictors were considered. The first baseline always predicts the most frequent outcome in the training set, typically a home win, reflecting the well known home advantage in football. The second baseline is a logistic regression model that uses only the Elo rating difference between the home and away teams. These baselines provide reference points for evaluating whether the Bayesian Network models and the inclusion of HMM derived momentum features offer improvements over simple, rating based approaches.

Conclusion

Our hybrid HMM BN model successfully integrates temporal dynamics (momentum) into static probabilistic models. The key finding is that simpler structures (Naive Bayes, Expert) outperformed complex data driven learning (HillClimb), highlighting that in noisy domains like football, complex structure learning

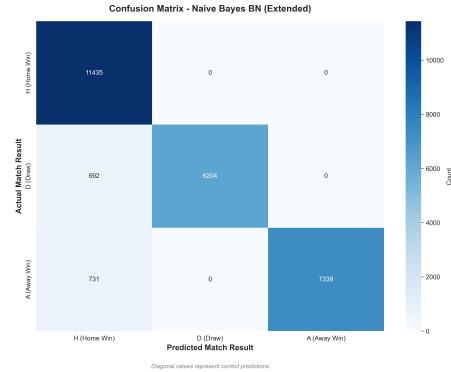


Figure 2: Confusion Matrix (Naive Bayes). Note the Draw Blindspot.

can lead to overfitting. The HMM features added value by capturing temporal patterns, but BN structure complexity did not translate to better predictions. Future work could explore dynamic Bayesian networks or ensemble methods while maintaining model simplicity.

Links to External Resources

Dataset: <https://www.kaggle.com/datasets/adamgbor/club-football-match-data-2000-2025>

Code Repository: <https://github.com/Fashad-Ahmed/Hybrid-Probabilistic-Model-for-Football-Match-Prediction>

References

- Constantinou, A. C., Fenton, N. E., & Neil, M. (2012). *pi-football: A Bayesian network model for forecasting Association Football match outcomes*. Knowledge-Based Systems, 36, 322–339.
- Rabiner, L. R. (1989). *A tutorial on hidden Markov models and selected applications in speech recognition*. Proceedings of the IEEE, 77(2), 257–286.
- Dixon, M. J., & Coles, S. G. (1997). *Modelling association football scores and inefficiencies in the fixed odds betting market*. Applied Statistics, 265–280.
- Koller, D., & Friedman, N. (2009). *Probabilistic Graphical Models: Principles and Techniques*. MIT Press.
- Bengio, Y., & Frasconi, P. (1995). *An input-output HMM architecture*. Advances in Neural Information Processing Systems (NIPS).