

Experimental Limits of Machine Learning on Near-Random Data: A Case Study of the Romanian Lottery System

Alexandru Dinu, Sara-Gabriela Modan, and Mădălin Frunzete(✉)

Faculty of Electronics, Telecommunications and Information Technology, National University of Science and Technology POLITEHNICA Bucharest, Bucharest, Romania
{alexandru.dinu2408,madalin.frunzete}@upb.ro, sara.gabriela.modan@stud.etti.upb.ro

Abstract. Lottery has attracted people's attention for a long time and the fascination and curiosity on how to beat the system are present today as they were at the beginning. This paper investigates if modern machine learning methods can be useful in extracting information and make predictions based on the Romanian 6/49 lottery historical data (more than 30 years of data and 2510 draws). First of all, we performed a statistical analysis based on the draw sequences and show that it is not statistically different compared to a uniform random process in terms of distribution, independence of data, and informational entropy. Secondly, based on the above-mentioned learnings, we evaluate four statistical and machine learning-based models: a frequency baseline, a Markov model, a multilayer perceptron, and a random forest classifier model. The training and testing data was carefully separated to avoid any leakage. The performance of these models is assessed against the expectation/hypothesis that they will deliver a performance close to random guessing. All models show almost identical hit rates, without any statistically significant difference compared to chance. The findings provide empirical evidence that when the data has information entropy closed to maximum (uniform distribution), no machine learning, as complex as it may be, is not able to achieve any systematic and predictable gains in the long run.

Keywords: Pseudo-random sequences · Lottery prediction · Machine learning limitations · Shannon entropy · Markov chains · Random Forest · Neural networks · Information theory

1 Introduction

Can artificial intelligence/machine learning methods help predict the lottery results? This is not a new question but still a very vivid and interesting one, not only for the general public but also for the researchers in the academia. Practically speaking, there is an entire (small) industry which focuses entirely on

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2027
O. Gervasi et al. (Eds.): ICCSA 2026 Workshops, LNCS 16759, pp. 358–371, 2027.
https://doi.org/10.1007/978-3-032-30518-3_23

finding a good-enough algorithm to "guess" next weeks lottery numbers (many of them, if not all). From a theoretical point of view, information theory can provide the answer: for a random process, no method can perform better than chance, no matter how elaborate and intricate that method is. The huge gap between the theoretical and the practical perspective is not usually closed empirically and that could be a reason why some people still struggle and work tirelessly to create a lottery-winning algorithm. Most published work either reports minor predictive gains without verifying in the first place that the data is truly random, or the alternative is that the scientific work dismisses the problem without correlating the rejection with the statistical properties of the specific dataset which is studied [1,2].

Romania's lottery system, Loto 6/49, is run by Loteria Romana and has been doing so continuously since 1993. Each draw selects 6 numbers without replacement from $\{1, 2, \ldots, 49\}$, producing a combinatorial space of $\binom{49}{6} = 13\,983\,816$ possible results/sequences. Of course, based on the physical safeguard mechanisms and procedures which are in place, one would expect that the draws follow a uniform distribution. Nonetheless, this assumption should be validated based on data and should not be taken for granted.

Previous work has investigated strategies based on frequency analysis [1], neural network predictors [2], first-order Markov models [3,4], and more recently LSTM-based approaches [5]. The results obtained across these studies are inconsistent, and sometimes positive findings are attributed to overfitting of test data which is not correctly separated in time [6]. Rigorous randomness test batteries such as the NIST Statistical Test Suite [7] and the Diehard tests [8] have been applied to lottery data before and generally support the hypothesis of randomness. However, these analyses are very rarely combined with a systematic machine learning evaluation applied on the same dataset.

This paper addresses that gap. We first verify, through multiple complementary statistical tests, that the Romania Loto 6/49 draw sequence is statistically indistinguishable from a uniform random process. We evaluate next four predictive models—a frequency baseline, a Markov chain, a Multilayer Perceptron, and a Random Forest classifier—against the theoretical random expectation, using strict temporal train/test separation to prevent data leakage. The central hypothesis is that no model will outperform random guessing, and we test this claim empirically rather than simply assuming it.

The current paper is organised as follows:

- Section 2 presents the statistical analysis of the lottery historical data
- Section 3 describes the experimental setup, machine learning models used for comparison, and results
- Section 4 discusses the broader implications and
- Section 5 concludes the paper and mentions the future investigation paths.

2 Statistical Analysis

The authors of this papers applied before statistical independence and uniformity testing via runs tests, autocorrelation analysis, and chi-square goodness-of-fit across diverse sets of sequential data, including chaotic dynamical systems and natural language corpora [9–12]. This provides a solid basis for the tests applied in this study.

The analyzed real dataset includes all official Romanian lottery system, Loto 6/49, winning sequences from August 8, 1993 to January 4, 2026, as published by Loteria Romana. Draw frequency increased from 1–2 per week in the early period to 2–3 per week after 2009. Years with fewer than 40 draws (1993 and 2026) are excluded from annual analyses, leaving 32 complete calendar years (1994–2025). After removing five draws which did not contain a complete sequence of 6 values, the cleaned dataset contains $n = 2510$ complete draws and 15060 (2510×6) individual number observations.

We establish the pseudo-random properties of the draw sequence before evaluating any predictive model, verifying three complementary properties: uniform distribution, near-maximum Shannon entropy, and serial independence of draws.

Uniform Distribution.
Figure 1 shows the observed experimental frequency of each number from 1 to 49, across all 2,510 draws, with counts ranging from 271 (number 31) to 354 (number 5) and a mean of 307.35. Each draw selects 6 numbers out of 49, every number has a theoretical probability of 6/49 of appearing in any given sequence. To verify that all 49 numbers are equally likely to be drawn—that is, that no number is systematically favoured or avoided—we apply the chi-square goodness-of-fit test:

$$\chi^2 = \sum_{i=1}^{49} \frac{(O_i - E_i)^2}{E_i} \tag{1}$$

where $E_i = N \cdot 6/49 = 307.35$. The result ($\chi^2 = 48.99$, $df = 48$, $p = 0.433$) provides no evidence against uniformity.

As an additional check, we look at the sum of the six drawn numbers for each draw. For example, if the numbers 3, 15, 22, 31, 40 and 47 are drawn, the sum is 158. Repeating this for all 2,510 draws gives us 2,510 such sums. The average of these sums, observed in the data, is 147.95, and they spread around this average with a standard deviation of 32.52—meaning most sums fall roughly between 115 and 180.

If the lottery system would be truly fair and all numbers are equally likely to be drawn, probability theory predicts that this average should be close to 150 and the spread close to 32.5. Our observed values of 147.95 and 32.52 match these predictions almost exactly, providing further confirmation that the draw process is uniform.

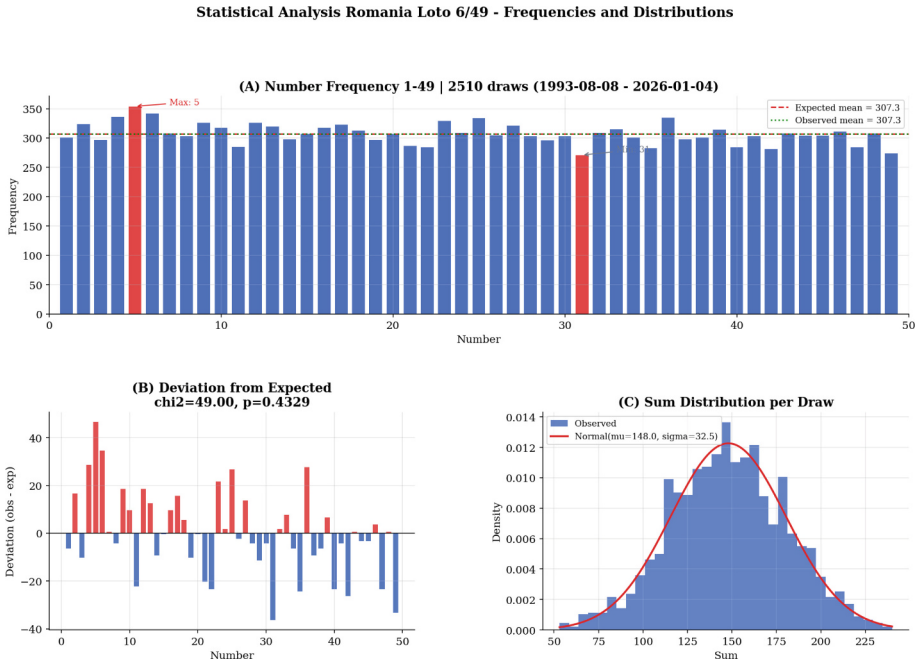

Fig. 1. (A) Observed frequency of each number (1–49) across 2510 draws (1994–2025). Red dashed line: expected frequency under uniformity (307.35); observed mean coincides at 307.35. Numbers 5 (maximum, 354 appearances) and 31 (minimum, 271 appearances) are annotated. (B) Deviation of each number' s frequency from the expected value of 307.35. (C) Distribution of draw sums with best-fitting normal curve.(Color figure online)

2.1 Shannon Entropy

Shannon entropy quantifies predictability directly. We compute the entropy based on 2:

$$H = -\sum_{i=1}^{49} \hat{p}_i \log_2 \hat{p}_i \tag{2}$$

and obtain $H = 5.6124$ bits compared to a theoretical maximum $H_{\max} = \log_2 49 = 5.6147$ bits, which gives an entropy ratio of 99.958%. Figure 2(D) confirms this is stable across all years, reducing the likelihood of temporal drift.

Independence of Draws.
A further important property to verify is that the draws are independent of one another—that is, the numbers drawn in one draw contain no information about the numbers that will appear in future draws. We test this in two complementary ways.

First, we apply the Wald-Wolfowitz runs test [13]. We take all 15,060 extracted numbers and convert them to a sequence of 0 s and 1 s: if a num-

ber is greater than 25 (the median) we write 1, otherwise 0. We then count the number of uninterrupted streaks of the same symbol—this concept is called *runs*. For example, in the sequence 1, 1, 0, 1, 0, 0 there are four runs: (1,1), (0), (1), (0,0). A run can be as short as a single value. If the sequence is truly random, the number of runs should be neither too high nor too low—too many runs would mean the sequence alternates too often, while too few would mean values cluster together more than expected. We observe $R = 7495$ runs against the expected value $\bar{R} = 7517.91$, computed from the counts of zeros and 1 s as $\bar{R} = 2n_0n_1/(n_0+n_1)+1$. This is a negligible difference ($Z = -0.374$, $p = 0.708$), consistent with a random sequence.

Second, we reshape all draws into a single sequence of 15,060 individual numbers in chronological order and measure whether any number tends to be correlated with the numbers appearing 1, 2, or up to 15 positions later in this sequence. This assessment is made by using the serial autocorrelation coefficient:

$$r_\ell = \frac{\sum_{j=1}^{n-\ell}(x_j - \bar{x})(x_{j+\ell} - \bar{x})}{\sum_{j=1}^{n}(x_j - \bar{x})^2} \tag{3}$$

All 15 coefficients depicted in Fig. 2 (A) are approximately zero—the largest is $|r_1| = 0.0085$, well within the 95% confidence bounds of $\pm 2/\sqrt{n} = \pm 0.0163$. This confirms that knowing which numbers appeared in past draws gives no useful information about future draws.

For each of the 49 numbers, we measure how many draws pass between two consecutive appearances—this is called the *gap* for that number. For example, if number 7 appears at draws 5, 12 and 31, its gaps are 7 and 19. Collecting all such gaps across all 49 numbers gives the gap sequence.

If the lottery is truly random and each number has an equal probability of $p = 6/49$ of being drawn at any given draw, then the waiting time between consecutive appearances of any number should follow a geometric distribution—that is, the gaps should decrease exponentially in frequency: short gaps are common, long gaps are rare. The observed gap sequence has mean $\bar{x}_g = 48.81$ and standard deviation $s_g = 45.32$, consistent with the theoretical expectation of $\lambda^{-1} = 49/6 \approx 8.17$ draws between appearances – Fig. 2 (C).

Next, a Kolmogorov-Smirnov test compares the observed gap distribution with the fitted exponential and yields $KS = 0.625$, $p < 0.001$, indicating a small but statistically detectable deviation. However, this deviation is fully expected and carries no predictive implications: since exactly 6 *distinct* numbers are drawn simultaneously at each draw, the same number cannot appear twice in the same draw. This introduces a slight regularity in the gap structure that causes a small deviation from a pure exponential—it is a mathematical consequence of the sampling mechanism and not a sign that the lottery system is predictable.

As a final check, we examine whether any two numbers tend to appear together in the same draw more often than would be expected by chance. For each of the $\binom{49}{2} = 1176$ possible pairs of numbers, we count how many times both numbers were drawn in the same draw—this is the co-occurrence count for

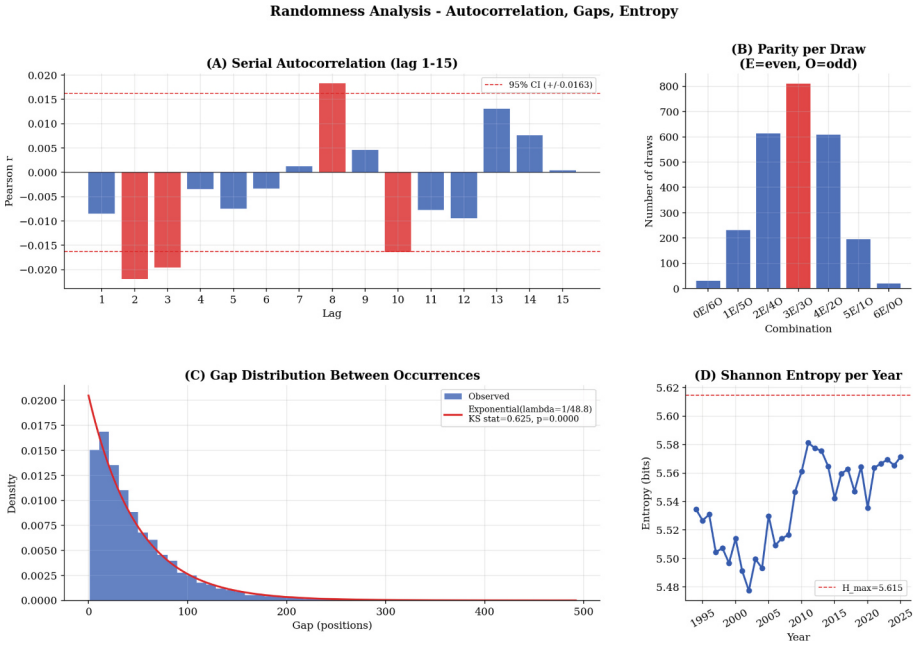

Fig. 2. Randomness analysis of the Romania Loto 6/49 draw sequence. (A) Serial autocorrelation for lags 1–15; dashed lines indicate 95% confidence bounds. (B) Parity distribution per draw; the most frequent combination is 3 even/3 odd (32.3%), consistent with combinatorial expectations. (C) Empirical gap distribution with best-fitting exponential curve. (D) Shannon entropy per year (1994–2025); dashed line indicates $H_{\max} = 5.615$ bits.

that pair. For example, if numbers 7 and 23 appeared together in 19 draws, their co-occurrence count is 19.

If the lottery would be fair and all numbers are independent of one another, each pair should appear together with equal probability at every draw. At each draw, exactly $\binom{6}{2} = 15$ pairs are formed among the 6 drawn numbers, out of a total of $\binom{49}{2} = 1176$ possible pairs. Over $N = 2510$ draws, the expected co-occurrence count for any specific pair is therefore:

$$E[C_{ij}] = N \cdot \frac{\binom{6}{2}}{\binom{49}{2}} = 2510 \cdot \frac{15}{1176} \approx 18.37 \tag{4}$$

The resulting 49×49 co-occurrence matrix is shown in Fig. 3. All observed counts fall close to 18.37, with no pair appearing together substantially more or less often than expected. No systematic pattern is visible—confirming that no number is statistically linked to any other number across the draw history.

Table 1 summarises the key statistical findings.

Figure 4 shows the maximum gap per number across all years; no systematic temporal pattern is visible, confirming stationarity.

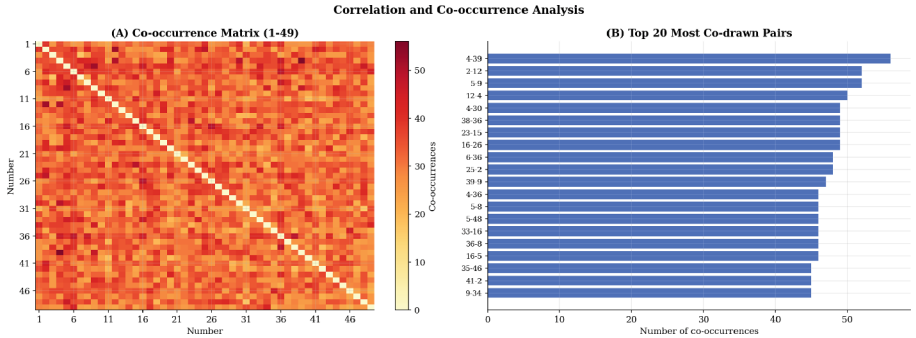

Fig. 3. Co-occurrence analysis. (A) Full 49×49 co-occurrence matrix; colour intensity indicates draw count per pair. No systematic pattern is visible. (B) Top 20 most frequently co-occurring pairs; counts are consistent with sampling variability around the expected value of approximately 18.4.

Table 1. Summary of statistical tests on Romania Loto 6/49 draw data (2,510 draws, 1994–2025).

Test	Null Hypothesis	Statistic	p-value	Decision
Chi-square	Uniform distribution	$\chi^2 = 48.99$	0.433	Not rejected
Shannon entropy	–	$H = 5.6124$ bits	–	99.96% of $H_{\max}$
Wald-Wolfowitz	Serial independence	$Z = -0.374$	0.708	Not rejected
Autocorr. ($\ell = 1$)	No linear dependence	$r = -0.0085$	–	Within 95% CI

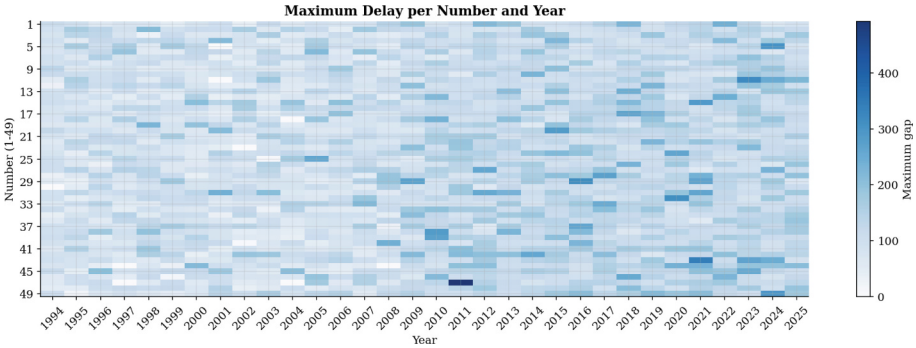

Fig. 4. Maximum gap per number and per year (1994–2025). Colour intensity represents the longest consecutive absence of each number within a given year. No systematic temporal pattern is visible, confirming stationarity of the process.

3 Experimental Setup and Results

Now that we established that the draw sequence is statistically the same as a uniform random process, we turn our attention to the central question of this

paper: can any machine learning model extract useful predictive structure from this data? In this section we describe how we set up the prediction task, which models we evaluate, and what results they achieve.

The main idea is very simple: we give each model access to the history of past draws and ask it to predict the 6 numbers that will be drawn next. We then compare how well each model performs against the simplest possible benchmark—picking 6 numbers completely at random. If a model is truly learning something useful from the data, it should predict correct numbers more often than pure random guessing. If it cannot, that is direct evidence that the data contains no learnable structure.

Each draw $d_t = \{n_1, \ldots, n_6\} \subset \{1, \ldots, 49\}$ is represented as a binary array $\mathbf{v}_t \in \{0,1\}^{49}$, where $v_{t,i} = 1$ if number i was drawn. For the ML models, a sliding window of $W = 10$ consecutive draws is used to construct the feature vectors $\mathbf{x}_t \in \mathbb{R}^{544}$, consisting of: the flattened one-hot encodings of draws $d_{t-W}, \ldots, d_{t-1}$ (dimension 490); five aggregate statistics over the window—normalised mean, standard deviation, sum, fraction of even numbers, and fraction of numbers ≤ 24 (dimension 5); and the normalised cumulative frequency vector up to time t (dimension 49). The target is $\mathbf{y}_t = \mathbf{v}_t$.

All models were implemented in Python using `scikit-learn` and `TensorFlow/Keras`. A strict temporal train/test split was used:: 80% for training (draws 1–2008, approximately 1994–2021) and 20% for testing (draws 2009–2510, approximately 2021–2025). Performance is measured by the hit count

$$h_t = |\hat{D}_t \cap D_t| \tag{5}$$

where $\hat{D}_t$ is the set of 6 predicted numbers and D_t the drawn set. The theoretical expectation under random prediction is

$$E[h] = 6 \cdot \frac{6}{49} = \frac{36}{49} \approx 0.7347 \tag{6}$$

derived from the hypergeometric distribution. This is the primary benchmark; real predictive power requires a mean hit rate significantly above this value.

3.1 Models

We evaluate five models of increasing complexity, ranging from non-parametric baselines to ensemble methods, to test whether predictive failure is model-specific or a general consequence of the data properties.

Model 0 – Random Baseline [1]. The random baseline selects 6 numbers uniformly at random without replacement, making no use of historical data. It converges in expectation to $E[h] \approx 0.7347$ and serves as the reference against which all other models are compared.

Model 1 – Historical Frequency [1]. This model predicts the 6 numbers with the highest cumulative frequency in all preceding draws:

$$\hat{D}_t = \underset{S \subset \{1,\ldots,49\},\, |S|=6}{\arg\max} \sum_{i \in S} F_{t-1}(i) \tag{7}$$

where $F_{t-1}(i)$ is the cumulative count of number i up to draw $t-1$. This embodies the common "hot numbers" heuristic. By the test period, cumulative frequencies are nearly uniform, so the model essentially selects a near-fixed set of slightly above-average numbers.

Model 2 – First-order Markov Chain [3,4]. This model estimates a 49×49 transition matrix $\mathbf{P}$ from training data, where P_{ij} is the empirical probability that number j appears given that i appeared in the preceding draw. Each candidate is scored as

$$s_j = 0.7 \cdot \sum_{i \in D_{t-1}} P_{ij} + 0.3 \cdot \hat{p}_j \tag{8}$$

with the second term regularising against sparse transitions using the empirical marginal frequency. Under true draw independence, $\mathbf{P}$ converges to a matrix with identical rows equal to $\mathbf{1}/49$, reducing the model to the setup described by model 1.

Model 3 – Multilayer Perceptron (MLP) [2]. The MLP maps $\mathbf{x}_t \in \mathbb{R}^{544}$ to the multi-label target $\mathbf{y}_t \in \{0,1\}^{49}$:

$$f_\theta(\mathbf{x}_t) = W_3 \cdot \mathrm{ReLU}(W_2 \cdot \mathrm{ReLU}(W_1 \mathbf{x}_t + \mathbf{b}_1) + \mathbf{b}_2) + \mathbf{b}_3 \tag{9}$$

The architecture uses three hidden layers (256, 128, 64 units, ReLU), Adam with L_2 regularisation ($\alpha = 10^{-4}$), and early stopping on a 10% validation split (patience 15 epochs). It is the most expressive model in the evaluation, and therefore it represents the most aggressive test of the hypothesis that no learnable structure exists.

Model 4 – Random Forest. The Random Forest model consists of 100 decision trees trained on bootstrap samples (max depth 10, min samples per leaf 5), implemented as a multi-output classifier with one binary tree per number. The predicted set consists of the 6 numbers with the highest predicted probability across all 49 classifiers:

$$\hat{D}_t = \underset{S \subset \{1,\ldots,49\},\, |S|=6}{\arg\max} \sum_{i \in S} \hat{P}(v_{t,i} = 1 \mid \mathbf{x}_t) \tag{10}$$

The ensemble structure provides inherent regularisation and makes the Random Forest a reliable detector of tabular structure when it exists [14].

3.2 Results

Table 2 summarises the quantitative results for all models based on the 500-draw test set. All models achieve mean hit rates within ± 0.09 deviation from the theoretical random expectation, and no model achieves more than 4 correct predictions on a single draw.

The Markov model records the highest mean hit rate and the only occurrence of 4 correct predictions (0.4% of draws). However, this can be due to the regularisation term in Eq. (8) creating a slight selection bias toward more frequent

Table 2. Model performance on the 500-draw test set. $\bar{h}$: mean hits per draw; s: standard deviation; Δ: difference from random baseline.

Model	$\bar{h}$	s	Δ	$P(h \geq 2)$	$P(h \geq 3)$
Theoretical random	0.7347	–	–	–	–
Random baseline	0.6880	0.742	0.000	13.6%	1.6%
Hist. frequency	0.7500	0.769	+0.062	16.2%	2.0%
Markov (order 1)	0.7740	0.771	+0.086	15.8%	2.0%
MLP neural net.	0.7080	0.789	+0.020	14.8%	3.0%
Random Forest	0.7360	0.742	+0.048	14.2%	1.8%

numbers. The MLP remains slightly below the random baseline—high-capacity models tend to concentrate on historically frequent numbers rather than sampling uniformly, which affects diversity. The Random Forest produces almost exactly at the theoretical expectation. This is arguably the most informative result: Random Forests are known to extract structure from tabular data when it exists, so failing to beat random here suggests there is little to no structure to exploit.

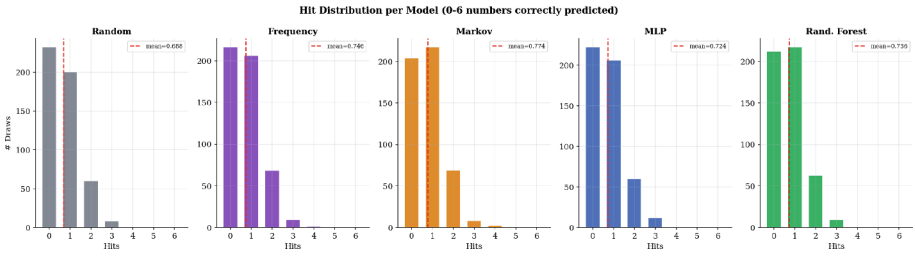

Fig. 5. Hit count distribution per model (500-draw test set).

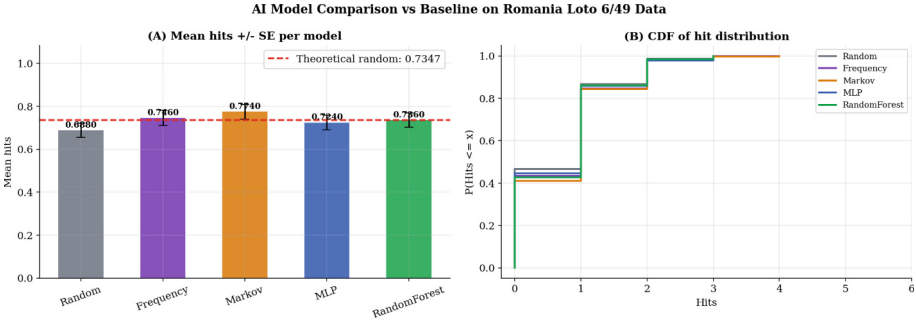

Fig. 6. Mean hit rates with standard error bars (A) and empirical CDFs (B).

Figure 5 shows the distribution of hit counts per model. Figure 6 compares mean hit rates and empirical CDFs across models. Figure 7 tracks rolling mean performance over the test period. Figure 8 breaks down results by hit category. Taken together, the four figures confirm that all models perform at the level of random guessing, with no sustained or structurally distinct above-random behaviour.

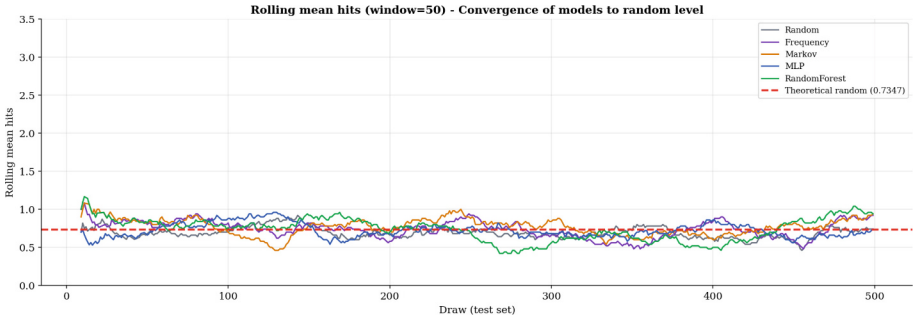

Fig. 7. Rolling mean hit rate (window = 50 draws) over the test period.

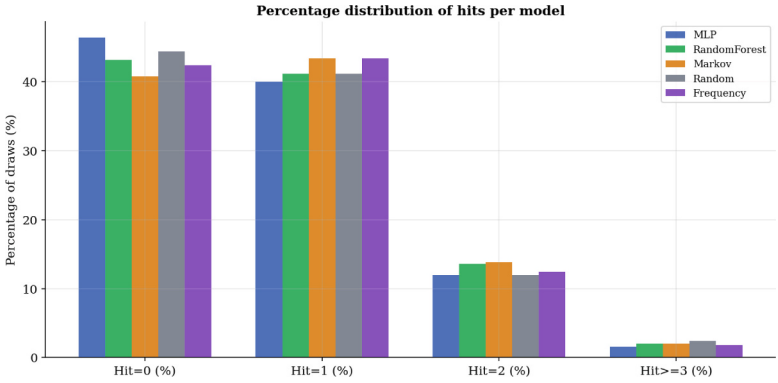

Fig. 8. Percentage of draws per hit category (0, 1, 2, $\geq$3) for all models.

4 Discussion

The lack of ability of all models tested in this paper to produce a good prediction follows directly from information theory basic results. Shannon's source coding theorem establishes that the minimum number of bits required to represent a

random variable is equal to its entropy [15]. A sequence with entropy near the theoretical maximum is essentially incompressible: no representation can encode it in fewer bits than the entropy value. As a direct consequence, any predictive model is searching for a compressed representation that retains information about future draws. However, when no such compression exists, the model is reduced to memorising the noise in the data. This explains why more complex models do not necessarily produce better predictions—the limiting factor is the absence of information in the data, not the model capacity or complexity.

An interesting feature of the results is that the most complex model (MLP) performs slightly worse than the simpler Markov chain. On pseudo-random data, a high-capacity model memorises various correlations in the training set that do not necessarily show general trends. Simpler models have fewer degrees of freedom to overfit and degrade to the random baseline. This has a practical implication beyond the lottery context: on pseudo-random data, model selection should always favour simplicity.

Generally speaking, these results add to the existing reasearch evidence which states that ML performance is limited by the statistical properties of the data, not just by the architecture and training procedure. In fields where data generation approximates a maximum-entropy process—financial data, quantum random noise generators, physical noise sources—the theoretical upper limit on predictability approaches random guessing regardless of the model used. Claims of predictive success on such data should be presented together with rigorous randomness testing, and performance should be compared with expectations based on solid theoretical grounds.

5 Conclusions

In the present paper, we have experimentally shown that machine learning algorithms cannot consistently perform better than random guessing. This conclusion is based on the historical Romanian lotter data, Loto 6/49, using 2,510 draws across 32 complete years (1994–2025). The draw sequence passes several complementary statistical tests for uniformity and statistical independence. In addition, the Shannon entropy for the data set is at 99.96% of its theoretical maximum. Four ML models—a historical frequency baseline, a first-order Markov chain, an MLP neural network, and a Random Forest classifier—were tested against the above-mentioned data set. They achieve mean hit rates consistent with the theoretical random expectation and no model manages to maintain consistent performance better than the average.

The results are not a failure of the specific algorithms put at test. They are a direct consequence of the information-theoretic properties of the data analyzed, and establish a practical upper bound on AI capability for this class of experiments. Future work could extend the analysis to lottery systems with different properties, or investigate whether alternative feature representations change the conclusions for more complex sequence models. The dataset and code can be made available to any interested researcher to support reproducibility of the results.

Acknowledgements. The authors acknowledge the use of publicly available historical data from Loteria Romana. This work was supported by UEFISCDI and National University of Science and Technology Politehnica Bucharest through projects PN-IV-P7-7.1-PTE-2024–0505 and UNSTPB-industry collaboration project 34808/19.09.2025.

Disclosure of Interests. The authors declare no competing interests. AI-assisted tools were used for language editing only; all scientific content and conclusions are the authors' own.

References

1. Haigh, J.: Taking Chances: Winning with Probability. Oxford University Press, Oxford (2003). https://doi.org/10.1093/oso/9780198526636.001.0001
2. Zhang, S., Liu, S., Xiong, J.: Why lottery ticket wins? a theoretical perspective of sample complexity on pruned neural networks. In: Advances in Neural Information Processing Systems (NeurIPS 2021), vol. 34 (2021)
3. Norris, J.R.: Markov Chains. Cambridge University Press, Cambridge (2012). https://doi.org/10.1017/CBO9780511810633
4. Farrell, L., Hartley, R., Lanot, G., Walker, I.: The demand for lotto: the role of conscious selection. J. Bus. Econom. Stat. **18**(2), 228–241 (2000). https://doi.org/10.2307/1392560
5. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural Comput. **9**(8), 1735–1780 (1997). https://doi.org/10.1162/neco.1997.9.8.1735
6. Wang, T.V., van Loon, R.J.D.P., van den Assem, M.J., van Dolder, D.: Number preferences in lotteries. Judgm. Decis. Mak. **11**(3), 243–259 (2016). https://doi.org/10.1017/S1930297500003089
7. Bassham, L., et al.: A statistical test suite for random and pseudorandom number generators for cryptographic applications. In: NIST Special Publication 800–22. National Institute of Standards and Technology, Gaithersburg, MD (2010). https://tsapps.nist.gov/publication/get_pdf.cfm?pub_id=906762
8. Marsaglia, G.: The Marsaglia Random Number CDROM Including the Diehard Battery of Tests of Randomness. Florida State University, Tallahassee, FL (1995)
9. Dinu, A.: From chaos to security: a comparative study of Lorenz and Rössler systems in cryptography. Cryptography **9**(3), 58 (2025). https://doi.org/10.3390/cryptography9030058
10. Dinu, A., Vlad, A., Hanu, B., Mitrea, A.: The statistical independence for words in printed Romanian language. In: 2020 13th International Conference on Communications (COMM), pp. 319–324, Bucharest, Romania (2020). https://doi.org/10.1109/COMM48946.2020.9142045
11. Dinu, A.: Singularity, observability, and independence: unveiling Lorenz's cryptographic potential. Mathematics **12**(18), 2798 (2024). https://doi.org/10.3390/math12182798
12. Dinu, A., Vlad, A.: Romanian printed language, statistical independence and the type II statistical error. In: 2021 International Conference on Speech Technology and Human-Computer Dialogue (SpeD), pp. 120–125 (2021). https://doi.org/10.1109/SpeD53181.2021.9587397
13. Wald, A., Wolfowitz, J.: On a test whether two samples are from the same population. Ann. Math. Stat. **11**(2), 147–162 (1940). https://doi.org/10.1214/aoms/1177731909

14. Breiman, L.: Random forests. Mach. Learn. **45**(1), 5–32 (2001). https://doi.org/10.1023/A:1010933404324
15. Cover, T.M., Thomas, J.A.: Elements of Information Theory, 2nd edn. Wiley (2006). https://doi.org/10.1002/047174882X