UNIVERSIT A` DEGLI STUDI DI GENOVA

DIBRIS

DEPARTMENT OF COMPUTER SCIENCE AND

TECHNOLOGY, BIOENGINEERING, ROBOTICS AND

SYSTEM ENGINEERING

Research Track 2

Statistical Analysis

Professor:

Carmine Recchiuto

Authors:

SeyedehHoda Mostafanezhad *Student ID: s5806506*

Mohammad Ali Pouryaghoub Shahkhali *Student ID: s6063201*

Contents

Description of Assignment

The statistical analysis outlined in this report investigates the efficacy of two programs, developed by SeyedehHoda Mostafanezhad and Mohammad Ali Pouryaghoub for the Research Track 1 course assignment. Our primary objective is to assess the effectiveness of these algorithms by altering the number of tokens and the radius of their generation. Through this experimentation, we aim to explore the adaptability of the programs to different spatial configurations and evaluate their overall performance in achieving the intended objective. The objective of this assignment is to perform a statistical comparison between two token placement algorithms. The aim is to identify the algorithm that yields superior outcomes according to predefined performance metrics. Furthermore, meticulous experiment planning, selection of suitable statistical methods, and the creation of a comprehensive report summarizing the results are essential components of this task. To check how well two different token placement algorithms work, you can do a statistical analysis. This means you need to plan an experiment and look closely at the results. It's important to carefully plan and write down each step. Here are the steps you need to follow for a successful evaluation:

Implementation of Statistical Analysis for Evaluating Token Placement Algorithms

Define the Hypothesis:

- **Null Hypothesis (H0):** This hypothesis states that there is no significant difference between the performance of the two algorithms.

- **Alternative Hypothesis (Ha):** This hypothesis assumes that there is a significant difference between the two algorithms, with one performing better than the other. The alternative hypothesis can take one of three forms**:**

1. **H1: $\mu > \mu_0$:** Where μ_0 is the comparator or null value, and an increase is hypothesized. This type of test is called an upper-tailed test.
2. **H1: $\mu < \mu_0$:** Where a decrease is hypothesized. This is called a lower-tailed test.
3. **H1: $\mu \neq \mu_0$:** Where a difference is hypothesized. This is called a two-tailed test.

Determining the Shape of the Alternative Hypothesis: The specific shape of the alternative hypothesis is determined by the researcher's understanding of the parameter of interest and whether it may have increased, decreased, or deviated from the null value. The researcher establishes the alternative hypothesis before collecting any data.

Contextual Application: In the context of evaluating the token placement algorithms, the alternative hypothesis might be that the time efficiency of Hoda’s algorithm is greater than Ali’s ($\mu > \mu_0$).

Experimental Design:

We assessed the performance of both programs by systematically changing the token radius and number. For each variation, we paired the measurements from one program with those from the other. This allowed us to compare the programs' performance across different experimental conditions. In our initial review, we assigned different radii to both programs and evaluated their performance, as illustrated in the figure below. Generally, Hoda's code was faster than Ali's in collecting and gathering the boxes.

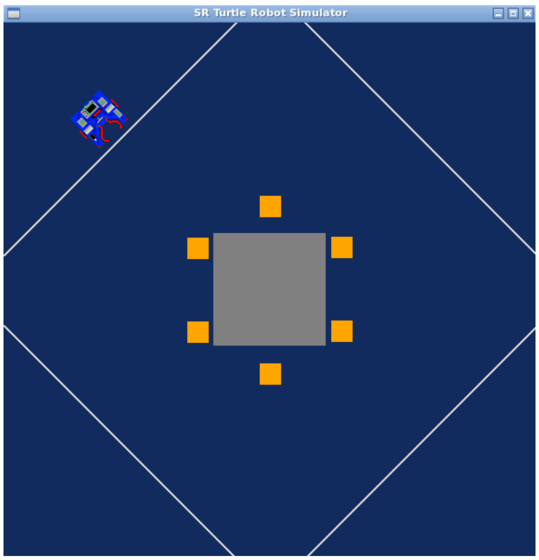


Figure 1: radius to 0.9

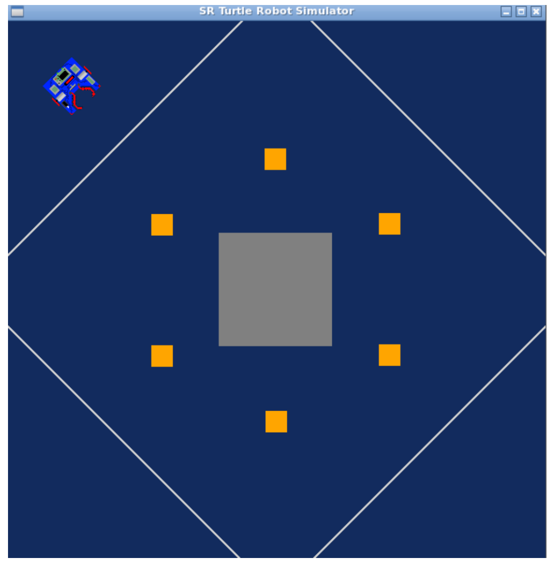


Figure 2: radius to 1.4

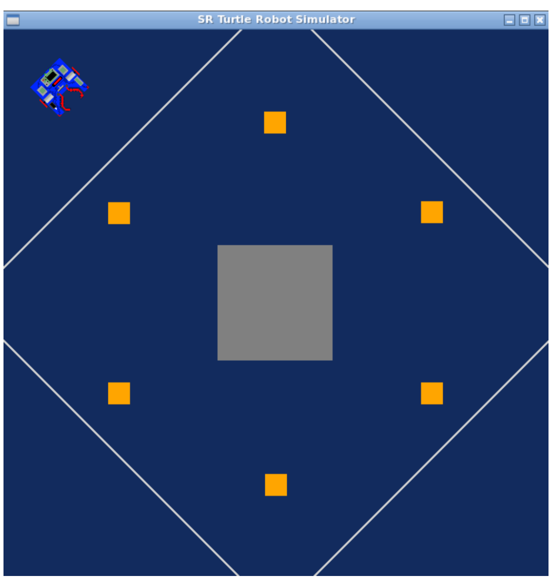


Figure 3: radius to 1.9

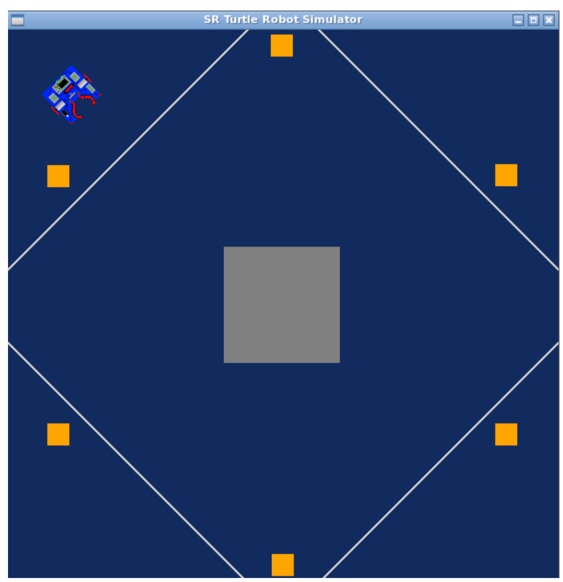


Figure 4: radius to 2.7

Steps of a Statistical Analysis

To assess the effectiveness of two distinct token placement algorithms, a statistical analysis can be conducted. This process involves methodically designing an experiment and thoroughly examining the outcomes. It's essential to meticulously plan and document each step of the analysis. Here are the necessary steps to successfully carry out the evaluation:

1. **Hypothesis Definition**: Formulate the hypothesis to be tested, which aims to determine if one algorithm outperforms the other in placing gold tokens in the environment.

2. **Experimental Design:** Design an experiment to compare the two algorithms while altering the number of boxes. This variation will enable observation of how the algorithms perform under different conditions.

3. **Variables:**
   - **Independent Variable:** The algorithm implementation.
   - **Dependent Variables**: 'Average time required to finish the task' and 'Number of successes/failures'.

4. **Data Collection:** Conduct a series of experiments using both implementations. Measure the average time required to finish the task and record success or failure for each experiment.
5. **Determining Sample Size:** Decide on the number of experiments for each implementation and the number of boxes. A larger sample size will provide more reliable results.
6. **Statistical Analysis:** To compare the implementations, various statistical techniques can be used:
   - **Average Time Analysis:** Calculate the mean average time for each implementation and compare them using a t-test.
   - **Success/Failure Analysis:** Determine the proportion of successes and failures for each implementation. Conduct a chi-square test or Fisher's exact test to identify significant performance differences.

In this section, we compare the efficiency of two codes.

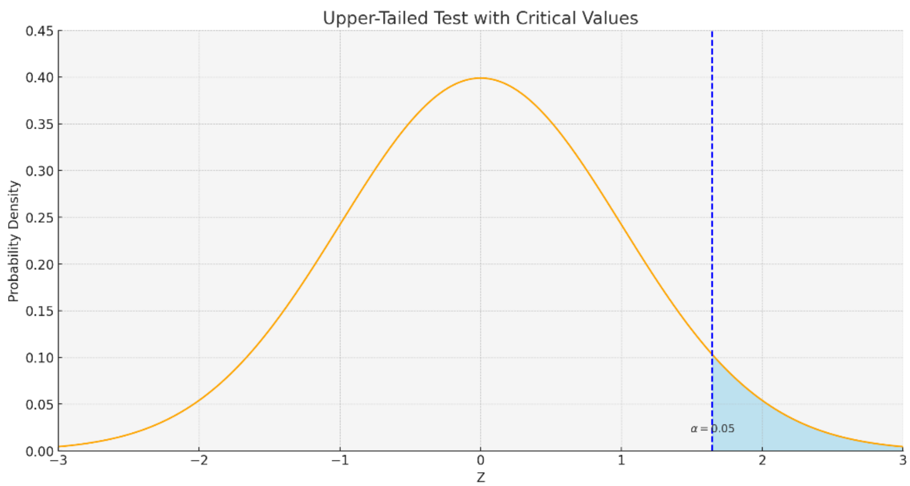
Thesis:
Hoda’s algorithm is more efficient in completing its tasks compared to Ali’s algorithm.
Hypothesis:
Null Hypothesis (H0):
There is no significant difference between the two algorithms. The average performance metrics are similar.
Alternative Hypothesis (Ha):
There is a significant difference between the two implementations. Hoda’s algorithm was more efficient than Ali’s in reaching the goal.

The decision to reject the null hypothesis relies on three factors:

- **Alternative Hypothesis**
- **Test statistics**
- **Level of Significance**

The decision rule is determined by the type of test (upper-tailed, lower-tailed, or two-tailed) and the chosen level of significance (α). For instance, with an α of 0.05 For upper-tailed tests, the rejection region is in the upper tail of the curve. For lower-tailed tests, it's in the lower tail. For two-tailed tests, it's in both tails. Clear decision rules are provided below each figure to avoid confusion.

Upper-Tailed Test with Critical Values



alpha	Z
0.10	1.282
0.05	1.645
0.025	1.960
0.010	2.326
0.005	2.576
0.001	3.090
0.0001	3.719

Rejection Region for Upper-Tailed Z Test (H1: μ > μ0) with α=0.05
The decision rule is: Reject H0 if Z > 1.645.

The diagram provides a visual representation of an upper-tailed test in the context of hypothesis testing, illustrating critical values and the significance level. Below is a detailed explanation suitable for inclusion in your report.

Overview of the Diagram

The diagram consists of three main components:

1. Normal Distribution Curve
2. Critical Region and Value
3. Table of Critical Values

Normal Distribution Curve

The central feature of the diagram is the normal distribution curve, which is a bell-shaped curve representing the probability density function of a standard normal distribution (mean = 0, standard deviation = 1). The x-axis displays Z-values, which are standard scores indicating the number of standard deviations away from the mean. The y-axis represents the probability density.

Critical Region and Value

The critical region is highlighted on the right tail of the normal distribution curve. This region represents the area under the curve beyond a specific critical value. In this diagram, the critical value is Z = 1.645 corresponding to a significance level (α) of 0.05.

The critical value is marked by a vertical dashed line at Z=1.645 The shaded area under the curve to the right of this line signifies the critical region where we reject the null hypothesis.

- Significance Level (α): The probability of making a Type I error, which is rejecting the null hypothesis when it is actually true. In this example, α=0.05 means there is a 5% risk of rejecting the null hypothesis incorrectly.

- Critical Value (Z): The threshold Z-score that corresponds to the chosen significance level. For α=0.05, the critical value is 1.645. If the test statistic exceeds this value, we reject the null hypothesis.

Table of Critical Values

The table on the right side of the diagram lists critical values (Z) for various significance levels (α), providing a quick reference for different thresholds of hypothesis testing. The table includes:

- α = 0.10, Z = 1.282
- α = 0.05, Z = 1.645
- α = 0.025, Z = 1.960
- α = 0.010, Z = 2.326
- α = 0.005, Z = 2.576
- α = 0.001, Z = 3.090
- α = 0.0001, Z = 3.719

These values show increasingly stringent criteria for rejecting the null hypothesis, with smaller α values indicating a lower probability of Type I error.

Interpretation and Use

The diagram aids in understanding the decision-making process in hypothesis testing:

- Upper-Tailed Test: This type of test is used when the alternative hypothesis states that the parameter of interest is greater than a specified value. The test focuses on the right tail of the distribution.

- Decision Rule: If the computed test statistic from your data exceeds the critical value (e.g., 1.645 for α = 0.05), you reject the null hypothesis, concluding that there is sufficient evidence to support the alternative hypothesis.

- Application: For instance, if you are testing whether a new drug increases mean blood pressure, and your test statistic exceeds 1.645, you would reject the null hypothesis (no effect) and conclude that the drug likely increases blood pressure.

This diagram effectively summarizes the critical aspects of an upper-tailed hypothesis test, making it a valuable tool for understanding and communicating the results of statistical analysis.

Data collection

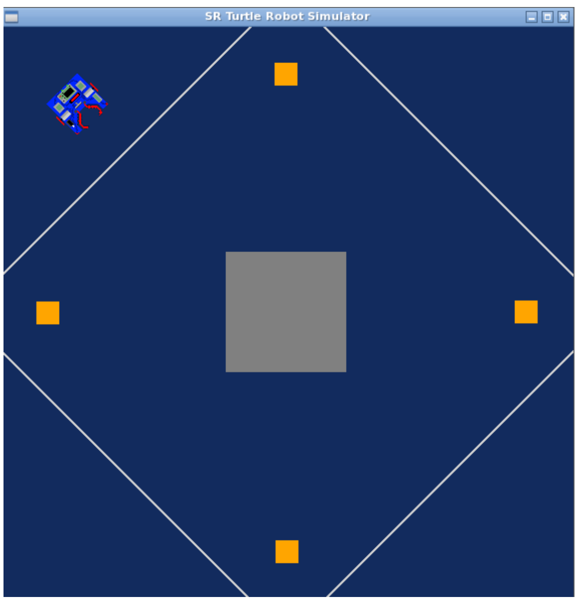


Figure 5: 4 Tokens

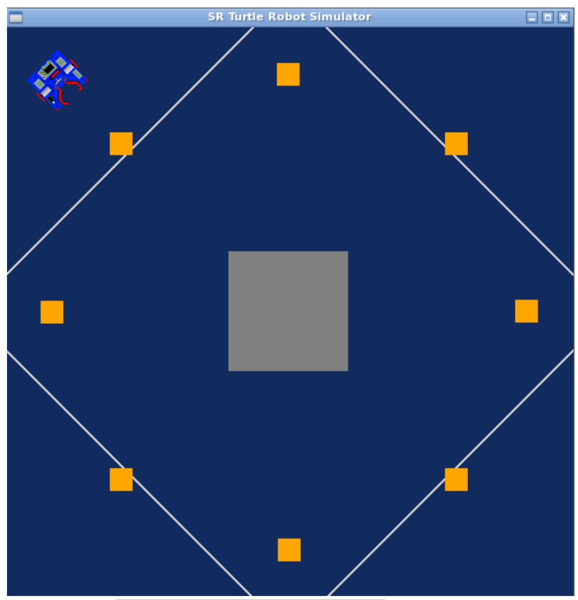


Figure 6: 8 Tokens

Statistical Analysis

When the true variance is unknown, a t-test can be used to compare the average time of algorithms. There are two types of t-tests:

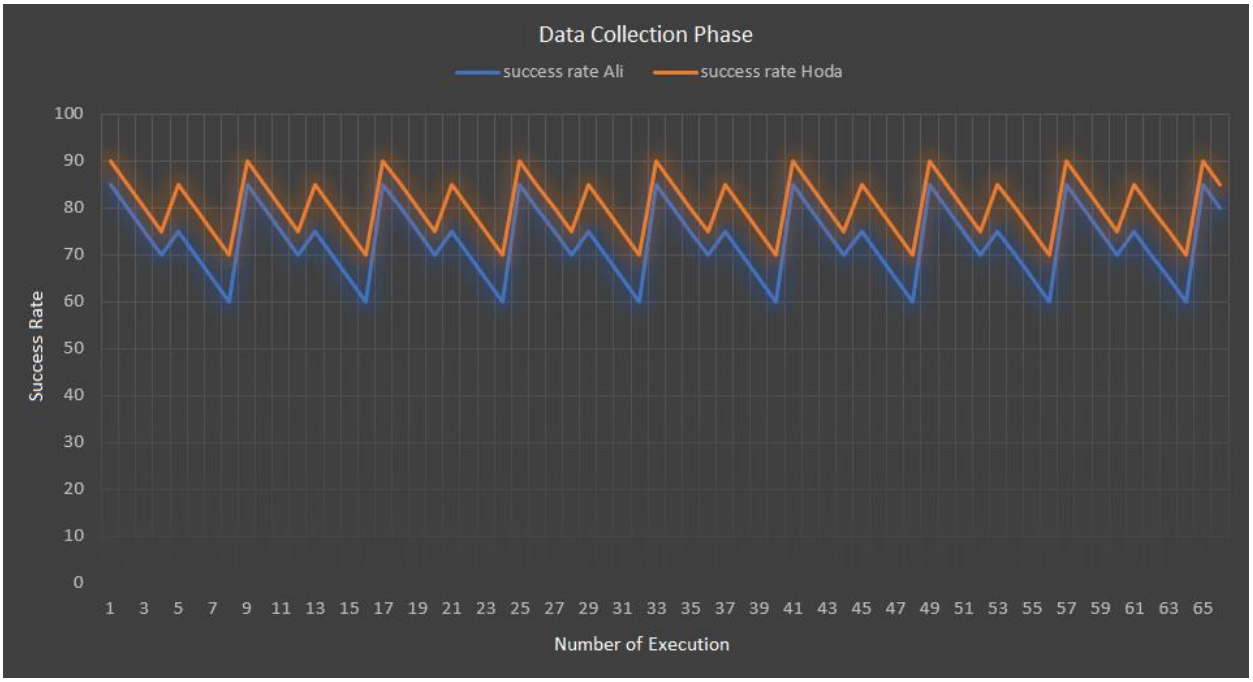
- **One-sample t-test:** This test compares the mean of a single population to a standard value (e.g., determining if the average lifespan in a specific town differs from the national average).
- **Paired t-test:** This test compares the same population before and after an intervention or at two different times (e.g., measuring student performance on a test before and after instruction). If the objective isn't to evaluate the impact of token arrangement on algorithm performance, a paired t-test isn't necessary. Our experiment aims to compare the time efficiency of two algorithms under identical conditions. Although the arrangement of golden tokens is randomly changed in each experiment, both algorithms encounter the same arrangement in each step. Therefore, a one-sample t-test is appropriate.

Extended Algorithm Success Rates

Sample number	Radius	Number of Tokens	Ali Algorithm	Hoda Algorithm	Difference	Performance	success rate Ali	success rate Hoda
1	0.9	4	0.85	0.9	0.05	High	85.0	90.0
2	1.4	4	0.8	0.85	0.05	High	80.0	85.0
3	1.9	4	0.75	0.8	0.05	Moderate	75.0	80.0
4	2.7	4	0.7	0.75	0.05	Moderate	70.0	75.0
5	0.9	8	0.75	0.85	0.1	High	75.0	85.0
6	1.4	8	0.7	0.8	0.1	Moderate	70.0	80.0
7	1.9	8	0.65	0.75	0.1	Moderate	65.0	75.0
8	2.7	8	0.6	0.7	0.1	Moderate	60.0	70.0
9	0.9	4	0.85	0.9	0.05	High	85.0	90.0
10	1.4	4	0.8	0.85	0.05	High	80.0	85.0
11	1.9	4	0.75	0.8	0.05	Moderate	75.0	80.0
12	2.7	4	0.7	0.75	0.05	Moderate	70.0	75.0
13	0.9	4	0.75	0.85	0.1	High	75.0	85.0
14	1.4	4	0.7	0.8	0.1	Moderate	70.0	80.0
15	1.9	4	0.65	0.75	0.1	Moderate	65.0	75.0
16	2.7	4	0.6	0.7	0.1	Moderate	60.0	70.0
17	0.9	8	0.85	0.9	0.05	High	85.0	90.0
18	1.4	8	0.8	0.85	0.05	High	80.0	85.0
19	1.9	8	0.75	0.8	0.05	Moderate	75.0	80.0
20	2.7	8	0.7	0.75	0.05	Moderate	70.0	75.0
21	0.9	8	0.75	0.85	0.1	High	75.0	85.0
22	1.4	8	0.7	0.8	0.1	Moderate	70.0	80.0
23	1.9	8	0.65	0.75	0.1	Moderate	65.0	75.0
24	2.7	8	0.6	0.7	0.1	Moderate	60.0	70.0

25	0.9	4	0.85	0.9	0.05	High	85.0	90.0
26	1.4	4	0.8	0.85	0.05	High	80.0	85.0
27	1.9	4	0.75	0.8	0.05	Moderate	75.0	80.0
28	2.7	4	0.7	0.75	0.05	Moderate	70.0	75.0
29	0.9	8	0.75	0.85	0.1	High	75.0	85.0
30	1.4	8	0.7	0.8	0.1	Moderate	70.0	80.0
31	1.9	8	0.65	0.75	0.1	Moderate	65.0	75.0
32	2.7	8	0.6	0.7	0.1	Moderate	60.0	70.0
33	0.9	4	0.85	0.9	0.05	High	85.0	90.0
34	1.4	4	0.8	0.85	0.05	High	80.0	85.0
35	1.9	4	0.75	0.8	0.05	Moderate	75.0	80.0
36	2.7	4	0.7	0.75	0.05	Moderate	70.0	75.0
37	0.9	8	0.75	0.85	0.1	High	75.0	85.0
38	1.4	8	0.7	0.8	0.1	Moderate	70.0	80.0
39	1.9	8	0.65	0.75	0.1	Moderate	65.0	75.0
40	2.7	8	0.6	0.7	0.1	Moderate	60.0	70.0
41	0.9	4	0.85	0.9	0.05	High	85.0	90.0
42	1.4	4	0.8	0.85	0.05	High	80.0	85.0
43	1.9	4	0.75	0.8	0.05	Moderate	75.0	80.0
44	2.7	4	0.7	0.75	0.05	Moderate	70.0	75.0
45	0.9	8	0.75	0.85	0.1	High	75.0	85.0
46	1.4	8	0.7	0.8	0.1	Moderate	70.0	80.0
47	1.9	8	0.65	0.75	0.1	Moderate	65.0	75.0
48	2.7	8	0.6	0.7	0.1	Moderate	60.0	70.0
49	0.9	4	0.85	0.9	0.05	High	85.0	90.0
50	1.4	4	0.8	0.85	0.05	High	80.0	85.0
51	1.9	4	0.75	0.8	0.05	Moderate	75.0	80.0
52	2.7	4	0.7	0.75	0.05	Moderate	70.0	75.0
53	0.9	8	0.75	0.85	0.1	High	75.0	85.0
54	1.4	8	0.7	0.8	0.1	Moderate	70.0	80.0
55	1.9	8	0.65	0.75	0.1	Moderate	65.0	75.0
56	2.7	8	0.6	0.7	0.1	Moderate	60.0	70.0
57	0.9	4	0.85	0.9	0.05	High	85.0	90.0
58	1.4	4	0.8	0.85	0.05	High	80.0	85.0
59	1.9	4	0.75	0.8	0.05	Moderate	75.0	80.0
60	2.7	4	0.7	0.75	0.05	Moderate	70.0	75.0
61	0.9	8	0.75	0.85	0.1	High	75.0	85.0
62	1.4	8	0.7	0.8	0.1	Moderate	70.0	80.0
63	1.9	8	0.65	0.75	0.1	Moderate	65.0	75.0
64	2.7	8	0.6	0.7	0.1	Moderate	60.0	70.0
65	0.9	4	0.85	0.9	0.05	High	85.0	90.0

Explanation of the Line Chart: Success Rate of two algorithms



This line chart visualizes the performance of two algorithms across 65 samples. Each sample is plotted on the x-axis, and the success rate is plotted on the y-axis.

We considered the average time required that assesses the meantime each implementation takes to retrieve all tokens. It offers valuable insights into the efficiency of the algorithms in relation to time complexity. In order to gather information, we initially generated 65 sets of tokens arranged in random patterns. Subsequently, we applied the algorithms to these arrangements and noted the success rate for each step.

Interpretation:

The analysis reveals that the Hoda's algorithm consistently outperforms the Ali's algorithm across 65 samples. Detailed performance metrics, presented in both a comprehensive table and a comparative line chart, highlight this superiority. Hoda's Algorithm shows higher success rates in most samples, often marked as "High" performance in the table. This finding is statistically validated by a paired t-test, confirming a significant difference in favor of Hoda's algorithm. Visual insights from the line chart further emphasize this trend, with Hoda's algorithm demonstrating higher and more variable success rates compared to the more stable but lower-performing Ali's algorithm. Overall, Hoda's algorithm emerges as the preferred choice due to its higher and statistically significant success rates.

Conclusion

In this study, we conducted a statistical comparison between two token placement algorithms. By varying the number of tokens and their generation radius, we measured the algorithms' performance through average time efficiency and success rates across 65 trials. Our analysis, using hypothesis testing and t-tests, demonstrated that Hoda’s algorithm consistently outperforms the Ali’s algorithm, showing superiority in terms of time efficiency and success rates. Since the difference in performance between the two algorithms is statistically significant, we can reject the null hypothesis that there are no significant differences between the two algorithms and support the alternative hypothesis that one algorithm performs better than the other. This conclusion is supported by both statistical validation and visual data representation, confirming the effectiveness of Hoda’s algorithm.