



RAISINS: Integrating R and AI for Agricultural Data Analysis

Mohammed Hisham M.
Statoberry LLP

Jithin Chandran
Statoberry LLP

Pratheesh P Gopinath 
Kerala Agricultural University

Abstract

RAISINS (R and AI Solutions for INferential Statistics) is a web-based statistical analysis platform designed to bridge the gap between complex inferential methods and the practical needs of students and researchers, particularly in the agricultural sciences. By integrating the computational power of R and Python with the adaptive capabilities of OpenAI's artificial intelligence APIs, RAISINS delivers an intuitive, code-free interface for performing advanced statistical analyses. Hosted on Posit Connect Cloud, the platform eliminates the need for local software installation, enabling seamless, browser-based access to a wide range of statistical tools. RAISINS simplifies model building, result interpretation, and reporting, making it especially valuable for non-statisticians and early-career researchers seeking reliable insights from agricultural data.

Keywords: Data Analysis, Agriculture Research, Design of Experiments, Statistical Genetics, Social Sciences, Open AI, Python, R.

1. Introduction

Statistical analysis serves as the backbone of scientific research, especially in disciplines like agriculture, where field experiments, survey data, and observational studies generate complex datasets. Despite the increasing emphasis on data-driven decision-making in agricultural research, many students, early-career researchers, and practitioners find themselves at a disadvantage when it comes to performing rigorous statistical analyses. The process is often hindered by a combination of factors: the steep learning curve associated with programming languages such as R and Python, the fragmented nature of statistical learning resources, and the lack of user-friendly tools tailored to agricultural research contexts.

While R and Python are powerful, flexible, and widely adopted in academic and professional settings, their command-line interfaces and syntax-heavy operations can be daunting for those without formal training in statistics or computer programming. Beginners frequently struggle with package selection, data cleaning, model specification, and interpretation of results. Furthermore, existing statistical software either lacks the flexibility and transparency of open-source tools or comes with licensing restrictions and high costs that limit accessibility in resource-constrained settings, particularly in developing countries.

In addition, agricultural research involves domain-specific needs such as handling experimental designs (e.g., CRD, RBD, split-plot), analyzing agronomic trials, processing survey data with complex sampling designs, and interpreting interactions between multiple environmental and genetic factors. Traditional statistical packages are not always well-adapted to these specific requirements, leaving researchers to either customize solutions on their own or consult specialists — both of which can delay or compromise the analytical workflow.

To bridge this gap, we developed RAISINS (R and AI Solutions for INferential Statistics) — a web-based platform designed to simplify and democratize statistical analysis for agricultural research. RAISINS integrates the power of R, Python, and AI-driven assistance within an intuitive interface that eliminates the need for code-based interaction. By abstracting complex syntax and offering guided, step-by-step workflows for data upload, analysis, visualization, and interpretation, the platform empowers users to focus on scientific inquiry rather than technical execution.

Built on open-source technologies and scalable cloud infrastructure, RAISINS provides modules for commonly used statistical methods — including descriptive analysis, hypothesis testing, regression models, design of experiments, reliability analysis, and more — while incorporating AI tools that assist in data summarization, model selection, and natural language explanations of results. This makes it an ideal solution for students, researchers, extension personnel, and educators engaged in agricultural and allied sciences.

In this paper, we present the design, features, and use-cases of RAISINS, and demonstrate how it addresses the unique challenges faced in agricultural data analysis by offering an accessible, robust, and intelligent statistical platform.

2. Materials and methods

2.1. Platform architecture and module validation

RAISINS is a modular cloud-hosted, web-based platform that integrates statistical computing,

artificial intelligence, and modern pedagogical design to serve agricultural researchers and students. Built using R, Python, and OpenAI APIs, RAISINS is deployed via **Posit Connect Cloud**, enabling browser-based access without requiring any local software installation. The platform is modular and each statistical tool within RAISINS is aligned with the theoretical underpinnings found in globally recognized academic literature.

All modules in RAISINS have been **peer-reviewed by domain experts**, including statisticians specializing in design of experiments, genetics, and social science research. This ensures both methodological rigor and instructional clarity.

Table 1 presents the major analytical domains addressed by RAISINS, the foundational academic references supporting each module domain, organized by area of application.

| Module Domain | Theoretical Foundation |
|------------------------------------|--|
| Design & Analysis of Experiments | Montgomery (2017) (Montgomery 2017); Gomez & Gomez (1984) (Gomez and Gomez 1984); Das & Giri (2018) (Das and Giri 2018); Dey (1986) (Dey 1986) |
| Statistical Genetics | Falconer & Mackay (1996) (Falconer and Mackay 1996); Lynch & Walsh (1998) (Lynch and Walsh 1998) |
| Social Science Research | Voss (2010) (Voss 2010); Babbie (2020) (Babbie 2020); Bryman (2016) (Bryman 2016) |
| Statistical Inference and Modeling | Casella & Berger (2002) (Casella and Berger 2002); Hogg, McKean & Craig (2018) (Hogg, McKean, and Craig 2018); Agresti (2013) (Agresti 2013) |
| Data Visualization & Reporting | Tufte (2001) (Tufte 2001); Cleveland (1993) (Cleveland 1993); Few (2009) (Few 2009) |

Table 1: Foundational literature supporting each RAISINS module domain, organized by area of application.

The algorithms and workflows integrated into RAISINS are grounded in established statistical theory, drawing extensively from authoritative texts and peer-reviewed literature. In cases where existing computational implementations were unavailable or insufficient, original algorithms were developed in adherence to rigorous theoretical principles. The platform’s design and interface have been iteratively refined based on user-centered feedback from researchers and educators, ensuring methodological soundness and usability consistent with global standards in scientific research and statistical education.

2.2. Packages and Libraries

RAISINS integrates a comprehensive ecosystem of open-source R and Python packages to deliver its core functionalities, encompassing statistical modeling, data visualization, artificial intelligence, and interactive web-based deployment. These packages support the platform’s

goal of offering high-level statistical capabilities through an accessible and intuitive interface, particularly beneficial for students and researchers in agricultural sciences.

The R-based architecture of RAISINS is built around the `shiny` framework (Chang, Cheng, Allaire, Xie, and McPherson 2024), which enables the development of dynamic, browser-accessible applications. Data manipulation tasks are performed using `dplyr` (Wickham, Francois, Henry, and Müller 2024), `tidyr`, and `data.table`, while data visualization is implemented through `ggplot2` (Wickham 2023), along with complementary packages such as `ggthemes`, `plotly` (Sievert 2023), and `ggpattern`. For statistical analysis relevant to agricultural experiments and general inference, the platform employs packages including `agricolae` (de Mendiburu 2020), `broom` (Robinson and Hayes 2024), `effectsize`, `lsr`, `car`, `phia`, and `FactoMineR` (Husson, Josse, Lê, and Mazet 2020). These libraries provide robust support for model fitting, hypothesis testing, effect size estimation, and multivariate analysis.

To enhance user interaction and responsiveness, RAISINS integrates several user interface and usability packages, such as `shinyWidgets`, `shinyalert`, `shinybusy`, `shinycustomloader`, `shinyBS`, `shinythemes`, and `colourpicker`. The presentation and export of analytical results are managed through tools like `rmarkdown`, `kableExtra`, and `formattable`, ensuring professional and structured output for reports and publications.

Python libraries are used to extend the platform’s analytical and AI-driven features. Packages such as `pandas` (The pandas development team 2024) and `numpy` (Harris, Millman, van der Walt, Gommers, Virtanen, Cournapeau, Wieser, Taylor, Berg, Smith *et al.* 2020) are employed for efficient data handling, while `matplotlib` and `seaborn` support graphical representations. For machine learning and statistical modeling, the platform uses `scikit-learn` (Pedregosa, Varoquaux, Gramfort, Michel, Thirion, Grisel, Blondel, Prettenhofer, Weiss, Dubourg *et al.* 2011) and `statsmodels` (Seabold and Perktold 2010). The `openai` package enables the integration of language models, facilitating intelligent assistance and automated content generation within the platform.

All R and Python packages used in RAISINS are well-maintained, community-supported, and widely recognized within academic and applied data science contexts. Each library is carefully selected based on its performance, stability, and relevance to agricultural research workflows. Modules are thoroughly reviewed by subject-matter experts to ensure methodological accuracy and pedagogical effectiveness, thereby maintaining the platform’s commitment to reproducibility, transparency, and educational value.

3. Illustrations

The Figure 1 illustrates the comprehensive architecture of the RAISINS platform, highlighting its integration of multiple programming languages and technologies. RAISINS combines the statistical computing power of R with the flexibility of Python and the advanced capabilities of AI through OpenAI models. The platform’s user interface is built using Shiny, providing an interactive and user-friendly experience, while the backend seamlessly coordinates R and Python processes. Hosting on Posit Connect Cloud ensures reliable accessibility and scalability for agricultural researchers. This integrated ecosystem enables efficient and robust inferential statistical analysis tailored specifically for agricultural research needs.

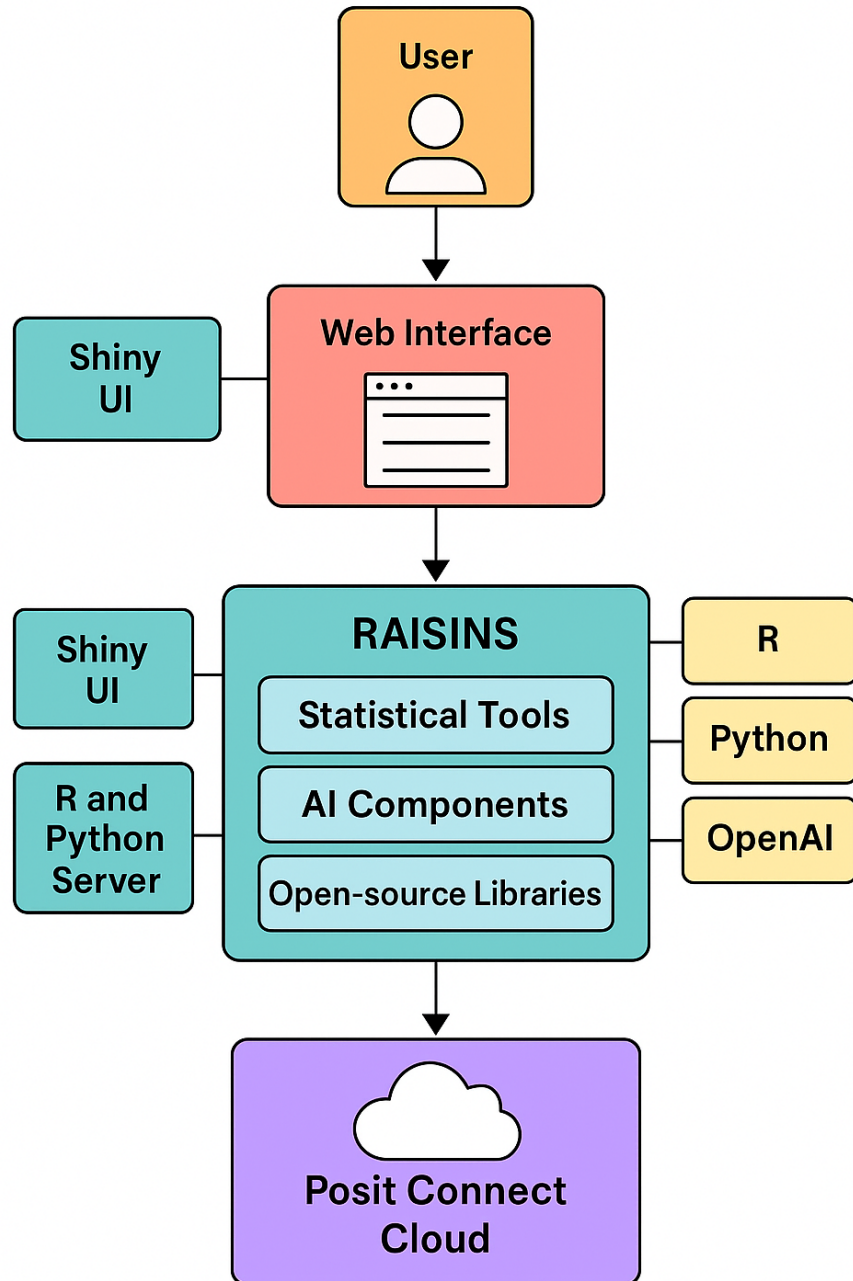


Figure 1: Architecture of the RAISINS Platform: Integration of R, Python, and AI Components through a Shiny-based Web Interface Hosted on Posit Connect Cloud.

4. Results

The RAISINS platform was developed to address the growing need for accessible, reliable, and AI-enhanced statistical tools tailored to agricultural research. By integrating familiar programming environments with user-friendly interfaces, it aims to bridge the gap between complex statistical methods and researchers' practical needs.

4.1. Use and Impact

The RAISINS platform has gained widespread adoption among postgraduate, and doctoral students, as well as faculty members across agricultural universities in India. Its intuitive design and web-based accessibility have made it an essential tool for academic coursework, thesis development, and research publications. By integrating AI-powered features, RAISINS significantly reduces the complexity of statistical output interpretation, making advanced inferential methods more approachable for users without extensive statistical backgrounds. This has empowered researchers to conduct rigorous data analysis with greater confidence and efficiency, thereby enhancing the overall quality and reproducibility of agricultural research.

Discussion

The RAISINS platform represents a significant advancement in making advanced statistical and AI-driven analytical tools accessible to the agricultural research community. By seamlessly integrating R, Python, and AI capabilities within an intuitive web interface, RAISINS addresses common barriers faced by researchers, such as the complexity of statistical software and interpretation of results. The platform's emphasis on user-friendliness and automation allows researchers, including those with limited statistical training, to perform rigorous data analysis with confidence. This democratization of statistical tools has the potential to improve the quality and reproducibility of agricultural research outputs. Moreover, the cloud-based deployment ensures scalability and continuous availability, enabling broader adoption across diverse educational and research institutions. The collaborative design and integration of multiple programming environments also open avenues for further enhancements, positioning RAISINS as a flexible and evolving resource in the field of agricultural data science.

Conclusion

The RAISINS platform marks an important step forward in democratizing access to advanced statistical and AI-powered analytical tools tailored specifically for agricultural research. By integrating the strengths of R and Python programming environments with artificial intelligence capabilities, RAISINS offers researchers an intuitive and comprehensive platform for conducting inferential statistics without requiring extensive technical expertise. Its web-based interface, supported by cloud hosting, ensures that users can access powerful data analysis tools anytime and from anywhere, thereby enhancing collaboration and research productivity. The platform's ability to simplify complex outputs and support diverse user groups—from students to seasoned faculty—has already contributed to improved data analysis quality and research rigor within agricultural universities. RAISINS not only supports robust statistical

inference but also fosters greater confidence among users in interpreting and applying results to their scientific inquiries. This positions RAISINS as a foundational resource in agricultural data science with substantial potential for continued growth and impact.

Future Work

Building on this foundation, future development efforts will focus on broadening RAISINS' analytical scope by incorporating modules for multivariate analysis and complex survey data, enabling users to tackle more sophisticated research designs. The introduction of voice-guided help and interactive tutorials aims to enhance user engagement and accessibility, particularly for those new to statistical analysis. Personalized user dashboards with automated report generation will streamline workflow management and facilitate documentation. Furthermore, refining the AI components by training on domain-specific agricultural datasets will improve the platform's ability to deliver context-sensitive insights and recommendations. These advancements will collectively strengthen RAISINS' role as an indispensable, evolving tool that meets the dynamic needs of the agricultural research community.

Acknowledgments

The authors gratefully acknowledge the support and encouragement provided by Kerala Agricultural University throughout the development of this work. We also extend our sincere thanks to the NABARD-KAU-WSU Agtech Launchpad for their invaluable incubation program and continued backing, which have been instrumental in advancing the RAISINS platform.

References

- Agresti A (2013). *Categorical Data Analysis*. 3rd edition. Wiley, Hoboken, NJ, USA.
- Babbie ER (2020). *The Practice of Social Research*. 15 edition. Cengage Learning.
- Bryman A (2016). *Social Research Methods*. 5 edition. Oxford University Press.
- Casella G, Berger RL (2002). *Statistical Inference*. 2 edition. Cengage Learning.
- Chang W, Cheng J, Allaire J, Xie Y, McPherson J (2024). *shiny: Web Application Framework for R*. R package version 1.8.0, URL <https://CRAN.R-project.org/package=shiny>.
- Cleveland WS (1993). *Visualizing Data*. Hobart Press.
- Das M, Giri N (2018). *Design and Analysis of Experiments*. 2nd edition. New Age International Publishers, New Delhi, India.
- de Mendiburu F (2020). *agricolae: Statistical Procedures for Agricultural Research*. R package version 1.3-5, URL <https://CRAN.R-project.org/package=agricolae>.

- Dey A (1986). *Theory of Block Designs*. Wiley Eastern Ltd, New Delhi, India.
- Falconer DS, Mackay TF (1996). *Introduction to Quantitative Genetics*. 4 edition. Longman.
- Few S (2009). *Now You See It: Simple Visualization Techniques for Quantitative Analysis*. Analytics Press, Berkeley, CA, USA.
- Gomez KA, Gomez AA (1984). *Statistical Procedures for Agricultural Research*. 2 edition. John Wiley & Sons.
- Harris CR, Millman KJ, van der Walt S, Gommers R, Virtanen P, Cournapeau D, Wieser E, Taylor J, Berg S, Smith NJ, *et al.* (2020). *NumPy: Fundamental Package for Scientific Computing with Python*. doi:10.1038/s41586-020-2649-2.
- Hogg RV, McKean J, Craig AT (2018). *Introduction to Mathematical Statistics*. 8 edition. Pearson.
- Husson F, Josse J, Lê S, Mazet J (2020). *FactoMineR: Multivariate Exploratory Data Analysis and Data Mining*. R package version 2.4, URL <https://CRAN.R-project.org/package=FactoMineR>.
- Lynch M, Walsh B (1998). *Genetics and Analysis of Quantitative Traits*. Sinauer Associates.
- Montgomery DC (2017). *Design and Analysis of Experiments*. 9 edition. John Wiley & Sons.
- Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, *et al.* (2011). *Scikit-learn: Machine Learning in Python*. URL <https://scikit-learn.org/>.
- Robinson D, Hayes A (2024). *broom: Convert Statistical Analysis Objects into Tidy Tibbles*. R package version 1.0.5, URL <https://CRAN.R-project.org/package=broom>.
- Seabold S, Perktold J (2010). *statsmodels: Statistical Modeling and Econometrics in Python*. Proceedings of the 9th Python in Science Conference, URL <https://www.statsmodels.org/>.
- Sievert C (2023). *plotly: Create Interactive Web Graphics via 'plotly.js'*. R package version 4.10.4, URL <https://CRAN.R-project.org/package=plotly>.
- The pandas development team (2024). *pandas: Powerful Python data analysis toolkit*. URL <https://pandas.pydata.org/>.
- Tufte ER (2001). *The Visual Display of Quantitative Information*. Graphics Press.
- Voss AD (2010). *Research Design and Statistical Analysis in Social Science*. Agricultural Research Press.
- Wickham H (2023). *ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics*. R package version 3.4.4, URL <https://CRAN.R-project.org/package=ggplot2>.
- Wickham H, Francois R, Henry L, Müller K (2024). *dplyr: A Grammar of Data Manipulation*. R package version 1.1.4, URL <https://CRAN.R-project.org/package=dplyr>.

Technical details

Due to the ongoing costs associated with hosting and maintaining the RAISINS platform, the authors have adopted a subscription-based model to ensure its sustainability and continued development. The live application is accessible to users at <https://www.raisins.live/>.

Publication Details

Submitted:

Accepted:

Affiliation:

Mohammed Hisham M.
Chief Technical Officer
Vellayani
Trivandrum India

Jithin Chandran
Data Analyst
Vellayani
Trivandrum India

Pratheesh P Gopinath
Department of Agricultural Statistics
Vellayani
Trivandrum India
E-mail: pratheesh.pg@kau.in
URL: <https://coavellayani.kau.in/people/sub-lt-dr-pratheesh-p-gopinath>