Internship Opportunity: Data Science & Training Set Optimization in Genomics

At **Heritable Ag**, we're building next-generation tools for data-driven breeding and plant improvement. As a fast-paced, Silicon Valley start-up, we combine cutting-edge statistical modeling, AI/ML, and domain expertise to advance agriculture with smarter, faster decisions. We're looking for a motivated intern to join our dynamic team and contribute to technology development involving machine learning, quantitative genetics, and high-performance computing.

🌱 Primary Project: Training Set Optimization

You will lead a core research project focused on algorithm-based optimization of training sets. Your work will include:

- Exploring how different optimization criteria perform in the context of multi-environment experiments. Applying and comparing different sampling strategies using both internal and public datasets
- Develop novel computational approaches for efficient use of small or medium-sized training sets to enhance plant prediction performance with diverse species

This is an exciting opportunity to contribute to research that directly supports better, faster, and more cost-effective plant prediction workflows.

🔧 Additional Projects/Tasks May Include:

- Develop and evaluate advanced ML methodologies for multi-trait genomic prediction
- Cleaning and preparing genomic and phenotypic datasets from ongoing customer projects
- Supporting automation of data pipelines in our infrastructure
- Monitoring recent research developments focused on developing and verifying new data analytics solutions

---

👤 Your Profile:

- Currently pursuing a PhD or have 3+ years of experience as a graduate student in a relevant field (e.g. quantitative genetics, biostatistics, data science)

- You are available to spend 3+ months in 2025 working at our offices in San Carlos, California
- A proven and demonstrable record of valuing, promoting, and advancing an inclusive and diverse work environment
- Strong interest in data analysis, with experience working on large and complex datasets
- Proficient in statistical programming (R and/or Python)
- Experience with HPC environments or Google Cloud Platform
- Strong problem-solving skills and a collaborative mindset
- Clear and effective communication skills

This internship offers the chance to work with a passionate, interdisciplinary team and contribute to impactful, applied research. You'll gain hands-on experience with data that drives decision-making in modern plant breeding and genomic applications.

Ready to apply or want to learn more? We'd love to hear from you!

Please send your application or any questions to:

✉ **Cathy Westhues** – [Senior Data Scientist]
Email: heritable-intern-2025@heritable.ag