

Sheetal Tewary

254-975-3253 | stewary2@illinois.edu | [/sheetal-t-21146585/](https://sheetal-t-21146585/) | [tewarysheetal](https://tewarysheetal.com)

EDUCATION

University of Illinois Urbana-Champaign

Illinois, USA

Masters in Statistics, Specialization in Data Science

Aug 2025–May 2027

Kalinga Institute of Industrial Technology

Odisha, India

Bachelor of Technology in Electronics and Instrumentation

Jul 2015–May 2019

PROJECTS

Tech Layoffs Analysis | *R, Statistical Modeling, Machine Learning*

Sep 2025–Jan 2026

- Analyzed 4,000+ tech layoff events (2020-2024) across 2,841 companies using LASSO logistic regression, achieving AUC of 0.60 for probability prediction and RMSE of 1.04 for magnitude modeling with 10-fold cross-validation in glmnet.
- Applied natural splines to model non-linear time trends and identified funding stage, industry sector, and country as key predictors of workforce reduction patterns through penalized regression analysis.

Concrete Strength Prediction | *SAS, Statistical Analysis, Classification*

Oct 2025–Dec 2025

- Built regression and classification models on 1,030 concrete samples to predict compressive strength from mix composition and curing age, achieving R^2 of 0.56 for mature concrete (≥ 100 days) and 95% accuracy for identifying high-strength mixes (≥ 50 MPa).
- Performed Ward's hierarchical clustering and ANOVA to segment mixtures into 4 distinct performance groups ($p < 0.0001$), applying Linear Discriminant Analysis (LDA) to classify samples by curing age with 65% accuracy.

True and Fake News Detection | *Python, Machine Learning*

Jun 2025

- Preprocessed and analyzed 20,000+ news articles using pandas and NLTK, implementing TF-IDF vectorization and training Logistic Regression and Naive Bayes classifiers in scikit-learn, achieving 97.9% accuracy with 5-fold cross-validation
- Engineered text features and compared model performance using precision-recall metrics, demonstrating robust classification across balanced test sets.

EXPERIENCE

Senior Data Consultant | *Informatica (Acquired by Salesforce)*

Jul 2022–Jul 2025

- Engineered scalable ETL pipelines processing 10+ billion records with optimized caching and load balancing, achieving 25% faster processing and 12% reduction in deployment downtime.
- Developed advanced data quality models and match algorithms using Address Doctor, improving data quality scores from 65 to 87 and reducing inconsistencies by 20%.
- Automated 10+ data onboarding workflows and implemented governance frameworks, enabling 2x throughput while maintaining regulatory compliance standards.

Data Engineer | *Accenture*

Jun 2019–Jun 2022

- Built scalable ETL pipelines integrating multi-source clinical data (Salesforce, Oracle) into star/Snowflake schema data warehouse, enabling advanced analytics on 10+ billion longitudinal records.
- Implemented slowly changing dimension (SCD Type 1/2) mappings and data masking protocols using Python, ensuring regulatory compliance and reducing data breach risks.
- Automated patient engagement workflows with real-time APIs and SQL-based notification systems, reducing manual effort by 40% and increasing patient engagement by 8%.

PUBLICATIONS

Design of CCII PID Controller for Glucose Blood Level Control using Genetic Algorithms, IEEE CICT-2018.

TECHNICAL SKILLS

Languages & Libraries: Python (pandas, NumPy, scikit-learn, NLTK, matplotlib), R (tidyverse, caret, glmnet, pROC), SAS, SQL

Statistical & ML Methods: Regression Analysis (Linear, Logistic, LASSO, Ridge), ANOVA, Hypothesis Testing, Time Series.

ML & Analytics : Supervised Learning (Classification, Regression), Unsupervised Learning (Clustering, PCA, LDA).

Data & Visualization: ggplot2, seaborn, Exploratory Data Analysis (EDA).

Data Engineering : ETL/ELT Pipelines, Data Quality & Profiling, MDM, Data Governance, Data Warehousing (Star/Snowflake Schema)

Tools & Platforms : Tableau, Power BI, Git, Jupyter, AWS, Databricks, Informatica Suite, Excel (VBA, Macros)