

# GEORGE SMITH

## DATA SCIENTIST

Berlin, Germany

[professorgeotheleo@outlook.com](mailto:professorgeotheleo@outlook.com)

+49 173 68 90 521

### PROFESSIONAL SUMMARY

Product-focused Data Scientist with 15+ years of experience applying machine learning, NLP, and experimentation to digital learning and SaaS products. Proven track record of translating complex data into actionable insights, improving user engagement, operational efficiency, and decision-making across academic-industry environments. Strong background in stakeholder collaboration, model evaluation, and end-to-end analytical ownership.

### KEY SKILLS

- **Applied Machine Learning:** Supervised & unsupervised learning, predictive modeling, feature engineering, model evaluation (AUC, precision/recall, RMSE)
- **Natural Language Processing (NLP):** Text classification, Transformers (Hugging Face), sentiment analysis, multilingual datasets
- **Experimentation & Causal Analysis:** A/B testing, controlled experiments, observational analysis, performance diagnostics
- **Product Analytics:** Recommendation systems, engagement metrics, funnel analysis, impact measurement
- **Data Quality & Validation:** Dataset construction, validation rules, error analysis, analytical robustness
- **Programming & Tools:** Python, scikit-learn, XGBoost, SQL, R (working knowledge)
- **Data Storytelling & Stakeholder Communication:** Translating analysis into product and business decisions

### WORK EXPERIENCE

#### Data Science Consultant / Freelance - Wall Street English

Mar 2021 - Jun 2025 Shanghai, China

- Designed and implemented an NLP classification pipeline using fine-tuned Transformer models (Hugging Face) to analyze customer feedback, automating review of 5,000+ monthly support tickets and reducing manual analysis by 20+ analyst-hours per week.
- Developed and evaluated a collaborative-filtering recommendation model in Python (scikit-learn), increasing average user session time by 15% in a controlled three-month A/B experiment.
- Conducted model evaluation and error analysis across NLP and recommendation systems, improving precision/recall and downstream decision reliability by ~12% for priority segments.
- Authored data-driven case studies on model interpretability for technical audiences; one publication exceeded 10,000 views and generated qualified consulting leads.

#### Data Science & Analytics Lead - Shixi International School, Suzhou International School

Oct 2018 - Mar 2021 Shanghai, & Suzhou China

- Built and validated predictive models (scikit-learn, XGBoost) for the IB Diploma Programme to forecast student exam outcomes using historical performance data (n=500+), achieving 88% accuracy in identifying at-risk learners.
- Led development of a large-scale multilingual text dataset (200,000+ translated sentence pairs), defining quality validation rules that improved dataset consistency from 85% to 99%.
- Diagnosed performance issues in a recommendation system through root-cause and feature-leakage analysis; retrained the model to improve precision by 12% for key learner segments.

## **Lead Data Scientist & Analytics Program Manager - The English Center/Argosy University**

Aug 2012 – Sep 2018 San Francisco Bay Area

- Owned end-to-end analytical strategy, developing gradient-boosting models (XGBoost) to predict outcomes and enable early interventions, contributing to a 7% reduction in failing grades.
- Designed analytical frameworks for evaluating student research performance and mentored 50+ candidates on statistical methods (R, SPSS); 95% of projects received top methodological scores.
- Built a skill-gap classification model for a B2B software training client, analyzing assessment data for 200+ engineers and achieving 92% alignment with manager evaluations.

## **Freelance - Senior Data Product Manager & Analytics Lead**

Aug 2012 – Dec 2016 Portland, OR

- Led analytics-driven development of 20+ adaptive learning modules, designing data collection strategies and A/B tests that improved learner retention metrics by an average of 25%.
- Built evaluation and feedback pipelines for early-stage learning products, reducing iteration cycles by 40% and increasing post-launch satisfaction scores by ~30%.

## **EDUCATION**

### **CALIFORNIA STATE UNIVERSITY, EAST BAY**

21 March 2014

MA – Teaching English to Speakers of Other Languages (TESOL)

7 September 2012

BA – English (Language & Discourse), Minor: Creative Writing