Ankit Kumar

Troisdorf, Germany | +49 176 87248106

chem.ak91@gmail.com | LinkedIn: ankit-kumar91- | GitHub: Ankit-kumar91

Professional Profile

Data Scientist with a strong scientific background in chemistry and analytical techniques. Skilled at integrating data-driven methodologies, including Python and machine learning, into R&D workflows. Experienced in managing multiple research projects, collaborating across departments, and independently executing experiments. Seeking to contribute to innovative pharmaceutical/biotech and chemical consulting projects in Germany.

Core Competencies

AI & ML: Machine Learning Algorithms, Deep Learning, Predictive Modeling, Pytorch, Scikit-Learn

Cheminformatics: RDKit, QSAR Modeling, Chemical Data Mining, Molecular Property Prediction

Data Science: Python Programming, Pandas, NumPy, Statistical Analysis, Data Preprocessing, Feature Engineering, Data Wrangling, Exploratory Data Analysis, Model Evaluation Metrics

R&D: Process Chemistry, Pharmaceutical Manufacturing, Yield Optimization, Scale-Up Development

Analytical Techniques: NMR Spectroscopy, FT-IR, UV-Vis, Fluorescence, HPLC, GC-MS, GPC, Mass Spectrometry, Chromatography, Spectroscopic Data Interpretation

Databases & Tools: SQL, Chemical Databases (PubChem, ChEMBL), Power BI, ChemDraw, SciFinder, PyMOL, Docker, Git, VS Code, Data Visualization

Professional Experience

AI Trainer — Expert & Reviewer   Mar 2025–Present
Outlier.ai | Troisdorf, Germany (Remote)

• Review, edit, and improve work submitted by peers to align with client requirements and quality standards

• Provide expert feedback on chemical data accuracy and scientific validity for AI training datasets

• Provide comprehensive feedback and ratings on peer submissions to support quality improvement

• Analyze the performance of AI models on chemistry-related tasks, identifying strengths, weaknesses, and areas for enhancement

Career Transition: AI & Machine Learning in Chemistry   Apr 2024–Feb 2025
Self-Directed Learning | Troisdorf, Germany

• Self-study in machine learning fundamentals, deep learning, and AI applications in chemistry

• Developed expertise in cheminformatics using RDKit for molecular property prediction and QSAR modeling

• Achieved B1 German language certification enhancing professional communication in German work environments

• NVIDIA-Certified Associate: Generative AI LLMs — Specialized training in large language models and AI applications

Data Science Trainee   Nov 2023–Mar 2024
neuefische GmbH — School for Digital Talent | Hamburg, Germany (Remote)

• Completed rigorous bootcamp covering machine learning algorithms, statistical modeling, and data science workflows

• Developed end-to-end ML projects including data collection, preprocessing, model training, and evaluation

• Gained proficiency in model performance metrics, cross-validation, and hyperparameter optimization

Full-Time Parenting   Jul 2022–Oct 2023
Troisdorf, Germany

Postdoctoral Research Associate   Dec 2020–Jun 2022
Technical University Dresden | Dresden, Germany

- Synthesized and analyzed carbon-based materials such as graphene nanoribbons and spin chains
- Developed scalable methodologies for hetero-structure graphene nanoribbons for optoelectronic applications
- Maintained lab records, managed equipment, and prepared comprehensive project reports

Senior Research Associate   Jan 2020–Oct 2020
Neogen Chemicals Limited | Vadodara, India

- Implemented process chemistry to develop scalable methods for mass production of pharmaceutical intermediates
- Synthesized 4-Methylphenol 4-IPMP with a 10% yield increase, reducing production costs and enhancing profitability
- Conducted cost analysis and managed projects for synthesizing 2,6-Difluorobenzyl bromide, achieving a 15% yield improvement
- Interpreted NMR, IR, and UV data, maintained laboratory records, and prepared reports

Doctor of Philosophy (Ph.D.)   Jul 2014–Dec 2019
Indian Institute of Technology Bombay | Mumbai, India

- Developed polyaromatic hydrocarbon based porphyrinoid macrocycles and metal complexes
- Investigated aromatic, antiaromatic, and non-aromatic behavior of macrocycles through experimental analysis and computational tools
- Established structure–property relationships linking molecular architecture with π-delocalization, electronic absorption, and redox behavior
- Strong experience with NMR, HR-MS, UV–Vis, Fluorescence, electrochemistry, and X-ray crystallography
- Published 13 peer-reviewed articles in high-impact journals

Education & Training

Doctor of Philosophy (Ph.D.) in Chemistry
Indian Institute of Technology Bombay, Mumbai, India | 2014–2019

Master of Science (M.Sc.) in Chemistry
CCS University Meerut, India | 2010–2012

Bachelor of Science (B.Sc.)
CCS University Meerut, India | 2007–2010

Key Data Science Projects

AqSolPrediction: Aqueous Solubility Prediction — GitHub | Live Demo
An end-to-end cheminformatics application for predicting aqueous solubility of chemical compounds using molecular descriptors and machine learning.

- Built ML pipeline using RDKit molecular descriptors for aqueous solubility prediction of chemical compounds
- Deployed interactive web app on Hugging Face Spaces with real-time SMILES input and prediction
- Implemented feature engineering with molecular fingerprints and physicochemical property calculations

Tech stack: Python, RDKit, Cheminformatics, Hugging Face, ML, Docker, FastAPI

AgronomiCare: Plant Disease Prediction — GitHub
A machine learning web application for agricultural disease prediction using classification algorithms with a complete data pipeline.

- Developed ML web application for agricultural disease prediction using classification algorithms
- Implemented complete data pipeline: data collection, cleaning, feature extraction, model training, deployment

Tech stack: Python, Machine Learning, Web App, Agriculture

Kickstarter Success Prediction — GitHub
Multi-model classification system comparing logistic regression, KNN, XGBoost, and Random Forest on real-world business data.

- Built multi-model classification system comparing logistic regression, KNN, XGBoost, and Random Forest
- Performed extensive feature engineering and data preprocessing on real-world business dataset

Tech stack: Python, XGBoost, Random Forest, Data Analysis

Research Publications

Author of 13 peer-reviewed scientific publications in prestigious international journals. Google Scholar: [Google Scholar]

Language Proficiency

German: Professional Working (B1 Certified) — Comfortable in professional environments

English: Full Professional Proficiency — Fluent in technical and business contexts

Hindi: Native Proficiency — Mother tongue

Additional Information

Work Authorization: Valid work permit for Germany; eligible to work for German companies

Availability: Immediate; open to positions in Germany, Cologne, Bonn, and DACH region

Mobility: Willing to travel for client projects and collaborative research initiatives