biotecnika
Your Bio Resource

Become a Data Scientist

DATA SCIENCE FOR BIOLOGISTS

Hands-On Internship

PROJECT | PAPER PUBLICATION | WORK EXPERIENCE

Extensive 120 Days
Hands-on Training With
3, 6 & 12 Months Project Work

Exclusively Crafted For B.Sc., M.Sc., B.Tech., M.Tech. &
PhD Biotech, Life Science, Chemistry + B.Pharm, M.Pharm Students

GET WORK EXPERIENCE

PAPER PUBLICATION ASSISTANCE

stores.biotecnika.org | Mail: info@biotecnika.org

biotecnika
Your Bio Resource

Embark on a transformative journey with the Data Science Hands-On Internship for Biologists: From Basics to Advanced with Projects & Publication! Tailored for biologists, this program offers expert mentorship, hands-on training, and practical experience with essential data science tools. From foundational skills to real-world projects and publications, you'll gain everything you need to thrive in a data-driven biology career. Don't miss this chance to innovate, discover, and grow—your future in data science starts here!

Extensive 120 Days Hands-on Training With 3,6 & 12 Months Project Work

WITH ONLINE + OFFLINE PROJECT

Session No.	Unit No. and Topic
UNIT 1: Introduction: Why Learn Bioinformatics?	
Day 1	Ice-breaking Session on Bioinformatics Global Tools and Techniques
Day 2	Scope & Opportunities in Bioinformatics: India & Abroad
Day 3	Understanding Key Terminology in Bioinformatics Databases
UNIT 2: Dicephering Molecules at the Sub-molecular level	
Day 4	Carbohydrates and Proteins **Demo on tools: Cheminformatics**
Day 5	Lipids and Nucleic acids **Demo on tools: ChemDraw & ChemSketch**

Program Module

UNIT 3: Bioanalytics: Understanding the patterns of life

Day 6	Primary Sequences – DNA, RNA, and Proteins **Demo on tools: NCBI, EMBL, DDBJ, and SBI**
Day 7	Secondary Structures – DNA, RNA, and Proteins **Demo on tools: GORV, Swiss model, I-TASSER, Phyre2**
Day 8	Tertiary Structures – Stability, folding, and misfolding **Demo on tools: Protein Data Bank (PDB), EMBL and SBI**
Day 9	Biochemical and Metabolic pathway analysis **Demo on tools: KEGG database and Cytoscape**

UNIT 4: Biocomputing- Computer-Aided Drug Design (CADD)

Day 10	Introduction, Drug discovery pipeline: Target identification and validation, Lead identification and optimization, Virtual screening **Demo: IMPPAT and Dr. Duke's Phytochemical and Ethnobotanical databases for Phytochemistry Research.**

Program Module

Day 11	ADMET Studies for the lead optimization. **Demo on tools:** SWISS-ADME, Analysis,and Interpretation Boiled-Egg Model
Day 12	Introduction to protein and modelling concepts, types, homology modelling, and the ab-initio method. **Demo on tools: PDB tools, Swiss ExPaSy**
Day 13	Constructing a Phylogenetic Tree- **Demo on tools: ClustalW, Clustal Omega**
Day 14	QSAR Studies and Interpretation of Docking **Demo on Molecular docking software:** Patchdock, Autodock, ClusPro web server
Day 15	Introduction to Molecular Dynamics and Simulation (MDS).

Unit 5: Biomimetics: Driving Deeper

Day 16	Pre-clinical (Bacterial, Fly, Plants, and Murine models), Human Clinical Trials for drug discovery, Functional genomics, and Comparative genomics. **Demo on the tool: Differential expression analysis on bacterial genome.**

Program Module

Day 17	Protein-protein interaction network, STRING, Network terminology, Pathway enrichment analysis **Demo on the tool: STRING and Cytoscape.**
Day 18	Human Genome project, Bacterial Genomics, Human Microbiome, and Epigenetics **Demo on database: Ensembl Bacteria, MicrobiomeDB**
Day 19	Sanger sequencing, Next Generation Sequencing: DNA Sequencing, RNA Sequencing. miRNA Analysis. **Demo on the tool: Databases: dbSNP, Clinvar, Genome browsers: UCSC and Ensembl**
Day 20	Basics of Genomics and Proteomics Introduction to NGS technologies: Overview and Workflow
Day 21	Next-Generation Sequencing (NGS) – DNA Sequencing **Demo on tools: FastQC, Trimmomatic, BWA**
Day 22	Hands-on NGS Practical Sessions Data Analysis and Visualization

Program Module

Day 23	NGS Data Preprocessing Using Python Scripts for Quality Filtering and Adapter Trimming

Unit 6: Penning Down: Publish or Perish

Day 24	Publication types, formats, and journals
Day 25	**Demo on the tool: Elsevier, Springer, Wiley, MDPI, and Bentham Sciences**
Day 26	Research made easy with Artificial Intelligence (AI & ML) **Demo on the tool: Applications of ChatGPT in BioResearch**
Day 27	Selection of the journal and indexing **Demo on the tool: Journal finder**

Unit 7: Proposal Writing and Grant Submission

Day 28	Submission, Revision, and re-submission process **Demo on tool: Editorial Manager**
Day 29	Structure and ingredients of proposal writing

Program Module

Day 30	Networking and funding from (govt./ private/ industrial partners)
Day 31	Portals for grant submission and prerequisites
Day 31	Generating primary data for proposal and funding

Unit 8: Coding for Biologist

Day 32	Introduction to coding for biologist
Day 33	Introduction to different User Interfaces • **Linux** • **Ubuntu** • **Windows** • **Mac**
Day 34	Location, Directory, path in computational terms. Basic configuration check of Computers.
Day 35	Cloud servers, Switching between different UIs, Basic Linux commands and their use

Program Module

Day 36	Linux in Bioinformatic analysis
Day 37	Getting Started with Python and R Python vs. R for Biological Data Analysis Installing Python, R, and Required Libraries
Day 38	Downloading and installation of python and IDEs
Day 39	Downloading and installation of R and RStudio
Day 40	Data Visualization Introduction to Data Visualization
Day 41	Creating Basic Plots with Matplotlib, Seaborn etc (Python)
Day 42	Creating Basic Plots with ggplot2 (R)
Day 43	Data Import and Manipulation Working with Biological Data Formats Basic Data Manipulation with Python and R
Day 44	Exploratory Data Analysis (EDA) Understanding Your Biological Data EDA Techniques in Python and R

Program Module

Day 45	**Introduction to Bio Python** What is Bio Python? Basic Bio Python Functions for Sequence Analysis
Day 46	**Sequence Feature Analysis** Identifying and Annotating Sequence Features Sequence Motif Search with Biopython
Day 47	**Protein Structure Analysis** Introduction to Protein Structure Analysis Using Biopython for Protein Structure Data
Day 48	**Phylogenetic Tree Analysis** Phylogenetic tree construction, visualization and interpretation.
Day 49	**Sequence Alignment** Introduction to Sequence Alignment Pairwise Sequence Alignment with Bio Python
Day 50	**Multiple Sequence Alignment** Multiple Sequence Alignment with Biopython Sequence Alignment Tools and Techniques
Day 51	**Clinical Data Preprocessing** Cleaning and Organizing Clinical Data Handling Missing Data

Program Module

Day 52	**Clinical Data Preprocessing Practicals Using Python** Biological examples
Day 53	**Clinical Data Preprocessing Practicals Using R** Biological examples
Day 54	**Survival Analysis** Introduction to Survival Analysis Kaplan-Meier Estimator and Cox Proportional Hazards Model in R
Day 55	**File Parsing and Data Retrieval** Reading and Writing FASTA Files Parsing GenBank Files
Day 56	Working with Sequence data and analysis using Biopython
Day 57	Biopython for Phylogenetic analysis
Day 58	**CADD data analysis using Python and BioPython** Working with molecular structures and visualization for drug designing Analyzing drug bioactivity data to screen potential drug candidate

Program Module

Day 59	Molecular structure visualization and analysis using python and biopython

UNIT-9: Introduction to AI and ML in Biology

Day 60	Overview of Artificial Intelligence (AI) and Machine Learning (ML)
Day 61	Introduction to Biology & Bioinformatics applications of AI/ML
Day 62	Importance and significance of AI/ML in biological research.
Day 63	Types of ML algorithms and their relevance in Biology & Bioinformatics
Day 64	Supervised learning algorithms and its types of ML algorithms
Day 29	Unsupervised learning algorithms and its types of ML algorithms

Program Module

Unit 10: Machine Learning in Bioinformatics

Day 66	Overview of supervised learning algorithms **Applications of in Biology & Bioinformatics (e.g., sequence classification, protein structure prediction)**
Day 67	Classification algorithms in details
Day 68	Regression algorithms in details
Day 69	AI/ML tools for CADD and its applications
Day 70	Dimensionality reduction techniques (e.g., Principal Component Analysis) and their applications.
Day 71	Unsupervised learning applications in Biology & Bioinformatics

Unit 11: Deep Learning in Bioinformatics

Day 72-73	Introduction to Deep Learning

Program Module

Day 74	Introduction to Neural Networks
Day 75	Convolutional Neural Networks (CNNs) for biological analysis
Day 76	Artificial Neural Network for biological data analysis
Day 77	Recurrent Neural Networks (RNNs) for sequence data analysis

Unit 12: Data Preprocessing in Bioinformatics

Day 78	Data cleaning techniques for biological data
Day 79	Handling missing data in bioinformatics datasets
Day 80	Data normalization transformation methods Feature selection dimensionality reduction techniques in bioinformatics
Day 81	Applications of AI/ML in Genomics.

Program Module

Day 82	Applications of AI/ML in Proteomics. Protein structure prediction techniques: AI-based modeling, AlphaFold etc
Day 83	Application of AI/ML in Drug Discovery.
Day 84	Applications of AI/ML in Multi-omics.
Day 85	Practical Sessions using GitHub repository Hands-on exercises using relevant tools and datasets.
Day 86	Practical Sessions using GitHub repository Hands-on exercises using relevant tools and datasets.
Day 87	Practical Sessions using GitHub repository Hands-on exercises using relevant tools and datasets.
Day 88	Practical Sessions using GitHub repository Hands-on exercises using relevant tools and datasets.
Day 89	Practical Sessions using GitHub repository Hands-on exercises using relevant tools and datasets.

Program Module

Day 90	**Introduction to Biostatistics** • **Overview of Biostatistics**: Definition, importance in public health and medicine, roles in research. • **Basic Concepts**: Variables (categorical, continuous), data types, populations vs. samples. **Descriptive vs. Inferential Statistics**: Explanation and examples.
Day 91	**Data Types and Measures of Central Tendency** • **Data Types Review**: Nominal, ordinal, interval, and ratio data. • **Central Tendency**: Mean, median, mode—when to use each. **Variation and Dispersion**: Range, variance, standard deviation, interquartile range.
Day 92	**Hypothesis Testing Basics** • **Null and Alternative Hypothesis**: Definitions and examples. • **Type I and Type II Errors**: Explanation of errors in hypothesis testing. **Significance Levels (α)**: Understanding p-values.

Program Module

Day 93	**Confidence Intervals** • **Understanding Confidence Intervals**: Meaning, interpretation, and significance. • **Confidence Interval for Mean**: How to compute and interpret. **Relationship to Hypothesis Testing**: Using CIs to assess statistical significance.
Day 94	**Sampling Methods** • **Sampling Techniques**: Simple random sampling, stratified, cluster, and systematic sampling. • **Sampling Bias and Its Impact**: Discuss selection bias, non-response bias. **Sample Size Determination**: Power analysis and sample size calculation
Day 95	**Study Design: Observational Studies** • **Types of Observational Studies**: Cross-sectional, case-control, cohort studies. • **Biases in Observational Studies**: Confounding, selection bias. **Advantages and Limitations**: Pros and cons of each study design.

Program Module

Day 96	**Study Design: Experimental Studies** • **Randomized Controlled Trials (RCTs)**: Key features and methodology. • **Blinding and Randomization**: Importance in reducing bias. **Ethical Considerations**: Ethics in clinical trials and human subject research.
Day 97	**Comparison of Means** • **Independent t-test**: Assumptions, application, and interpretation. • **Paired t-test**: When to use it, examples, and calculations. **One-Way ANOVA**: Introduction to analysis of variance.
Day 98	**Non-Parametric Tests** • **When to Use Non-Parametric Tests**: For non-normal data or ordinal variables. • **Mann-Whitney U Test**, **Wilcoxon Signed-Rank Test**, **Kruskal-Wallis Test**. **Chi-Square Test**: For categorical data.

Program Module

Day 99	**Regression Analysis** • **Linear Regression**: Simple linear regression, model interpretation, assumptions. • **Multiple Regression**: Multiple predictors, multicollinearity, model selection. **Model Diagnostics**: Checking residuals, goodness-of-fit.
Day 100	**Logistic Regression** • **Introduction to Logistic Regression**: Binary outcomes, odds ratio. • **Model Building**: Interpretation of coefficients, model selection, and diagnostics. **Assumptions of Logistic Regression**: Linearity, absence of multicollinearity.
Day 101	**Survival Analysis** • **Introduction to Survival Data**: Censoring, survival time, event vs. non-event. • **Kaplan-Meier Estimator**: Estimating survival probabilities. **Cox Proportional Hazards Model**: Hazard ratio, multivariable survival models.
Day 102	**Correlation and Association** • **Pearson Correlation**: Measures of linear correlation. • **Spearman's Rank Correlation**: Non-parametric measure of association. • **Chi-Square for Association**: In categorical data analysis.

Program Module

Day 103	**Whole Genome Sequencing (WGS): Applications and Analysis** • WGS Workflow: Data Generation, Mapping, and Variant Calling • Applications in Clinical Diagnostics, Population Genomics, and Evolutionary Studies
Day 104	**Whole Exome Sequencing (WES): Enrichment and Disease Focus** • WES Workflow: Target Enrichment, Alignment, and Variant Analysis • Industrial Applications: Rare Disease Identification and Precision Medicine
Day 105	**RNA-Seq for Transcriptomics: Expression and Beyond** • RNA-Seq Workflow: Mapping, Quantification, and Differential Gene Expression Analysis • Applications in Biomarker Discovery, Cancer Diagnostics, and Drug Response
Day 106	**Single-Cell RNA-Seq (scRNA-Seq): Decoding Cell Heterogeneity** • scRNA-Seq Analysis: Dimensionality Reduction, Clustering, and Cell-Type Identification • Applications in Developmental Biology and Precision Oncology

Program Module

Day 107

Long-Read Sequencing: PacBio and Nanopore Technologies

- Benefits of Long-Read Sequencing for Structural Variants and Complex Genomes
- Case Studies: Genome Assembly for Novel Species and Clinical Genomics

Day 108

Metagenomics: Decoding Microbial Diversity

- NGS in Metagenomics: Tools like QIIME2, HUMAnN, and MEGAN
- Applications in Industrial Biotechnology, Environmental Monitoring, and Human Microbiomes

Day 109

Epigenomics Using NGS: DNA Methylation and Chromatin Accessibility

- Techniques: Bisulfite Sequencing, ATAC-Seq, and ChIP-Seq
- Applications in Cancer Epigenetics and Developmental Biology

Day 110

Cancer Genomics: Somatic vs Germline Analysis

- Identifying Somatic Mutations, Copy Number Variations, and Gene Fusions
- Case Studies: Liquid Biopsy, Circulating Tumor DNA (ctDNA), and Precision Oncology

Program Module

Day 111	**Multi-Omics Data Integration Using NGS** • Integrating Genomics, Transcriptomics, and Epigenomics for Biomarker Discovery • Industrial Applications: Systems Biology for Disease Mechanisms
Day 112	**Whole Genome and Transcriptome Sequencing for Rare Diseases** • Combining WGS and RNA-Seq for Disease Mechanisms • Applications in Clinical Genetics and Personalized Healthcare
Day 113	**Machine Learning and Artificial Intelligence for NGS** • ML Algorithms for Variant Prioritization, Gene Expression Prediction, and Biomarker Identification • Tools and Case Studies: Deep Learning Applications in Genomic Data
Day 114	**Advanced Visualization and Reporting for NGS Data** • Tools: Circos, UCSC Genome Browser, IGV, and Shiny Apps • Generating Industry-Standard Reports for Clinical and Industrial Applications

Program Module

Day 115	**Advanced Error Correction and Quality Enhancement in NGS Data** • Challenges in NGS Data: Sequencing errors, alignment inaccuracies, and bias in variant calling. • Error Correction Methods: K-mer based approaches for raw read correction. Probabilistic models (e.g., Bayesian approaches) for base calling.
Day 116	Doubt solving session
Day 117	Doubt solving session
Day 118	Open discussion
Day 119	Test and Project instructions
Day 120	Project topic discussion

PROJECT TOPIC LIST

1. Novel lead optimization and presentation against HMPV/SARS-CoV-2/Mpox/Dengue Virus: An In Silico Drug designing and discovery approach

2. Targeted antioxidants against mutant proteins responsible in Neurodegenerative disorders like ALS/FALS, Alzheimer's, Huntington's and Multiple Sclerosis using Structure-based Drug designing approach

3. Identifying novel mutations in different complex disorders using AI-ML-based tools and designing the mutant protein model with quality and stability evaluation

4. Different types of Cancer metabolomic studies to identify the affected cascade and to come up with personalized medicine solutions using CADD and Data science

5. Analysing the Pharmacokinetic and Pharmacodynamic properties of Novel Drug molecules using AI-ML-based tools to help patients manage major complex disorders in a better way

PROJECT TOPIC LIST

6 Drug repurposing against Respiratory disorders with pocketome and cavitome analysis using Computer Aided Drug Designing (CADD) and Machine Learning (ML)

7 Natural Lead molecules virtual screening library preparation, Docking against desired receptor and post dock analysis using AI-ML based tools and web servers

8 Next-Generation Sequencing for Comprehensive Pathogen Variant Detection, Tracking, and Surveillance in Infectious Diseases

9 Genomic Profiling with NGS for Tailored Treatment Approaches in Neurological Disorders: A Personalized Medicine Approach

10 Exploring the Tumor Microenvironment with Single-Cell RNA-Seq and AI to Discover Novel Cancer Biomarkers and Therapeutic Pathways

11 Predicting and Mitigating CRISPR Off-Target Effects with Single-Cell RNA-Seq and AI Models to Improve Gene Editing Precision

PROJECT TOPIC LIST

12 Identification of Disease-Associated Genetic Variants via GWAS Analysis

13 Comparative Genomics Analysis of Viral Genomes

14 AI-Driven Identification of Epigenetic Biomarkers for Aging

15 NLP for Automated Extraction of Drug-Disease Relationships from Literature

16 Reinforcement Learning for Optimizing CRISPR-Cas9 Guide RNA Design

17 Graph Neural Networks (GNNs) for Protein-Ligand Interaction Prediction

18 AI-Powered Quality Control for Single-Cell RNA-Seq Data

19 Deep Learning for Predicting RNA-Protein Interactions Using Raw Sequence Data

20 Prediction of HIV Drug Resistance Mutations - Use supervised learning to identify mutations in HIV that confer drug resistance based on sequence data

PROJECT TOPIC LIST

21 Uncovering Novel RNA Editing Sites in Neurodegenerative Diseases Using Machine Learning

22 Developing AI-Based Tools for Predicting Protein Function from Sequence Data

23 Identification of potential drug candidates by CADD-SBDD based approach

24 Integrating Deep Learning and Ensemble Learning for Enhanced High-Throughput Screening Data Analysis in Drug Discovery

25 AI/ML model of bioinformatics for predicting antimicrobial resistance in bacteria

26 Identifying disease-associated genetic variants using AI/ML based tool

27 Multi-Omic data integration for identifying disease-associated variations

28 Neural Network Model to Predict Splicing Events from Genomic Sequences

29 AI-based Biomarker Discovery for Early Cancer Detection - Utilize deep learning models to identify potential biomarkers for early-stage cancer from genomic and proteomic data

biotecnika
Your Bio Resource

PROJECT TOPIC LIST

30 Prioritizing Pathogenic Variants in *BRCA1/2* Using AlphaMissense Scores & ClinVar Annotations for mutant protein modelling of Breast/Ovarian Cancer

31 Identifying Novel CFTR Mutations in Cystic Fibrosis Patients via AlphaMissense-Driven Reanalysis of Negative Genomic Screens

32 Exploiting Cryo-EM Structures of Mutant CFTR for Correctors in Cystic Fibrosis Therapy

33 Structure-Guided Design of Covalent Inhibitors for KRAS G12D in Pancreatic Cancer

34 Targeting the PCSK9-LDLR Interaction Interface with Peptidomimetics for Hypercholesterolemia

35 Characterizing LMNA Variants in Dilated Cardiomyopathy: Integrating AlphaMissense Scores with ClinVar Pathogenicity and dbSNP Allele Frequencies

36 Designing Selective LOXL2 Catalytic Domain Inhibitors Using Active-Site Mapping for Fibrotic Disorders

37 Rational Design of SOD1 Dimer Interface Stabilizers Using Crystal Structures for ALS (Amyotrophic Lateral Sclerosis)

PROJECT TOPIC LIST

38 Discovery of Rare SOD1 Mutations in Familial ALS by Combining AlphaMissense Predictions with dbSNP Minor Allele Frequency Filtering

39 Designing Selective HER2 Kinase Inhibitors Using Co-Crystallography Data for Metastatic Breast Cancer

40 Effect of Mutations on Protein-Protein Interactions: A Case Study on TP53 and Its Variants

41 Analyzing the Impact of Spike Protein Mutations on SARS-CoV-2 Infectivity and Immune Escape

42 Predicting Alzheimer's Disease Risk Using Polygenic Scores (PGS)

43 QSAR Modeling for Bioactivity Prediction of Kinase Inhibitors

44 Deep Generative Models for Molecule Design (VAE, GAN, RL)

45 Rare Disease Diagnosis Prediction from Symptoms (Multi-label Classification)

COURSE FLOW - Training → Test → Project Interview → Project begins → Report Submission

About the Instructor

Ms. Nilofer K Shaikh , PhD

Biotecnika CRO scientist/ ALML & Bioinformatics Specialist

With a strong background in big data analysis using computational approaches in cancer omics data, Ms. Nilofer K Shaikh brings a wealth of experience from MIT ADT University. Her expertise spans cancer research, drug design, molecular dynamics simulation, data mining, and various omics technologies. Proficient in Python, R, and computational methodologies, she has a deep understanding of genomics, metabolomics, proteomics, transcriptomics, pharmacogenomics, and AI for cancer treatment. Her skillset also includes machine learning, MySQL database management, and natural language processing (NLP

About the Instructor

1.
2.

Mr. Prodyot Banerjee

Biotecnika CRO scientist/ CADD, Bioinformatics & Genomics Specialist

Prodyot Banerjee is a seasoned professional in Computer-Aided Drug Designing, Bioinformatics Analysis, and Genomics, boasting rich experience from institutions like CSIR-IGIB, CSIR-CLRI, IIT Madras, and Delhi Technological University.

With an M.Tech in Bioinformatics from Delhi Technological University, Prodyot has excelled in research and development roles, presenting his work at prestigious venues like IIT Kharagpur. His research is published in esteemed journals such as IEEE and Frontiers in Pharmacology, with more underway. Prodyot's GATE 2019 qualification from IIT Madras underscores his dedication to both academic excellence and professional growth. With a proven track record and relentless pursuit of knowledge, he is a valuable asset in bioinformatics, genomics, and computer-aided drug design endeavors.

About the Instructor

Dr. Elamathi Natrajan

Biotecnika CRO scientist/ ALML & Bioinformatics Specialist

Elamathi Natarajan is a dedicated bioinformatician with a robust background in computational biology, data analysis, and genomics. Holding a Doctorate in Bioinformatics from Dr. A.P.J Abdul Kalam Technical University and an MBA in Information Systems Management, she has made significant contributions to the field through both research and teaching.

She has served as an Assistant Professor and Head of Department (HOD) In-Charge at Kalinga University, Raipur, where she excelled in lecturing, research, and departmental management. At Biotecnika Info Labs Pvt Ltd, Bangalore, she played a key role in academic support, enhancing student success through coaching and program development.

Elamathi's expertise includes developing bioinformatics pipelines, conducting quality assessments, and applying machine learning algorithms to genomics data. Recognized for her work, including a Senior Research Fellowship from the Indian Council of Medical Research (ICMR), she continues to drive innovation in bioinformatics and is seeking a new challenge to further advance scientific discoveries.

About the Instructor

Ms. Shubhi Singh

Biotecnika CRO scientist/ Bioinformatics & Computational Drug Design Specialist

Ms. Shubhi Singh is a dedicated researcher and educator having completed her Ph.D. in Biotechnology from SRM Institute of Science and Technology, bringing a strong interdisciplinary background in both computational and experimental research.

She is proficient in molecular docking, ADMET prediction, protein–ligand interaction analysis, and bioinformatics workflows for target identification and drug design. Her expertise includes integrating in silico modeling with wet-lab techniques such as qPCR-based gene expression profiling, microbial culture under stress conditions, and biofilm inhibition assays.

Ms. Singh is also experienced in academic teaching and laboratory supervision, fostering scientific skills and critical thinking among students. She is dedicated to advancing research in antimicrobial resistance and therapeutic development through innovative, data-driven approaches.

About the Instructor

1. **Dr. Bhupender**

Biotecnika CRO scientist/AI/ML, Bioinformatics & Computational Drug Design Specialist

Dr Bhupender Singh is currently working as a Scientist/Trainer at Biotecnika info Labs. He holds a PhD degree specialising in Bioinformatics from Lovely Professional University. During his PhD, he grew as a dynamic researcher with experience as Teaching Assistant and has published his thesis work on Machine Learning and in-silico Drug Discovery in peer-reviewed journals.
With extensive experience at CSIR-Institute of Microbial Technology (IMTECH) as a Project Scientist and Senior Project Associate, he has proven his skills in ML/AI, Cheminformatics, in-silico Drug Discovery, Protein-sequence and Structural Analysis, Database development, supervision and organisation of the Biological-sampling events, gamification of scientific concepts for awareness of children and others. Apart from this, he has also taken scientific administrative roles at Lovely Professional University contributing to PhD thesis submission processes and Plagiarism verification.

Dr Bhupender has published in several peer-reviewed journals and has more than 700 citations. He has active collaborations with reputed National and International Scientists, Academicians, and Researchers. Further, he is also serving as a Reviewer in peer-reviewed scientific journals of Bioinformatics and Computational Biology for more than four years.
Dr Bhupender has joined Biotecnika with an objective to share his learning and experience among growing professionals.

Leading Biopharma Companies Hiring Biologists for Data Science Abroad

Illumina:

Illumina leverages AI and genomic data analytics to improve precision medicine, cancer genomics, and agricultural biotechnology.

Google (DeepMind, Verily)

Google has strong AI/ML research arms (DeepMind) and a dedicated healthcare subsidiary (Verily) actively involved in bioinformatics, drug discovery, and personalized medicine.

Microsoft (Microsoft Research)

Microsoft Research conducts cutting-edge research in AI, including areas relevant to biomedicine. They also offer cloud services and tools for life sciences.

Johnson & Johnson

J&J integrates AI and machine learning across pharmaceutical R&D, medical devices, and healthcare services. They use predictive analytics to streamline clinical trials, optimize patient recruitment, and personalize treatments.

Pfizer

An American multinational pharmaceutical and biotechnology company. They hire bioinformaticians for roles in drug discovery, clinical research, and vaccines development.

23andMe/AncestryDNA

These companies specialize in consumer genomics and utilize data science for genetic research and personalized health insights.

GSK

GSK invests heavily in AI for vaccine development and drug discovery. Their AI-powered analytics platforms enhance clinical trial design, monitor real-world data, and optimize supply chains for pharmaceutical manufacturing.

AstraZeneca

AstraZeneca employs AI and data science for precision medicine, improving drug discovery efficiency and clinical trial accuracy. They integrate machine learning tools to analyze complex disease pathways and predict patient responses to treatments.

Flatiron Health

Flatiron integrates real-world data analytics into oncology research, bridging the gap between clinical trials and routine patient care. They structure data from Electronic Health Records (EHRs) to accelerate cancer treatment insights and improve healthcare delivery outcomes.

Recursion Pharmaceuticals

Recursion combines machine learning, bioinformatics, and automation to analyze massive biological datasets. Their AI models identify new drug targets and repurpose existing drugs for rare diseases, significantly shortening the discovery timeline.

Interdisciplinary Nature

Data Science offers biologists unique chances to integrate their expertise with computational tools. Combining biology, statistics, and machine learning allows them to tackle challenges like genomic analysis, ecosystem modeling, and drug interaction predictions. This synergy fosters innovative solutions in healthcare, agriculture, and environmental science, positioning biologists as essential contributors to data-driven research.

Research Opportunities

Data Science provides biologists with research opportunities for complex questions with computational precision. From bioinformatics to biodiversity modeling, options such as drug discovery through machine learning and omics integration for personalized medicine enhance understanding and drive innovations in healthcare, agriculture, and sustainability.

Data Science is transforming biology, enabling biologists to analyze complex datasets and drive innovation. From genomic data in personalized medicine to enhancing crop yields, this integration has created exciting career paths. Biologists possess essential domain expertise, allowing them to utilize data science effectively. In healthcare, environmental science, and drug discovery, the ability to interpret biological data is crucial for research and technological advancement.

CAREER SCOPE IN DATA SCIENCE FOR BIOLOGISTS

Toll Free - 1800-1200-1818

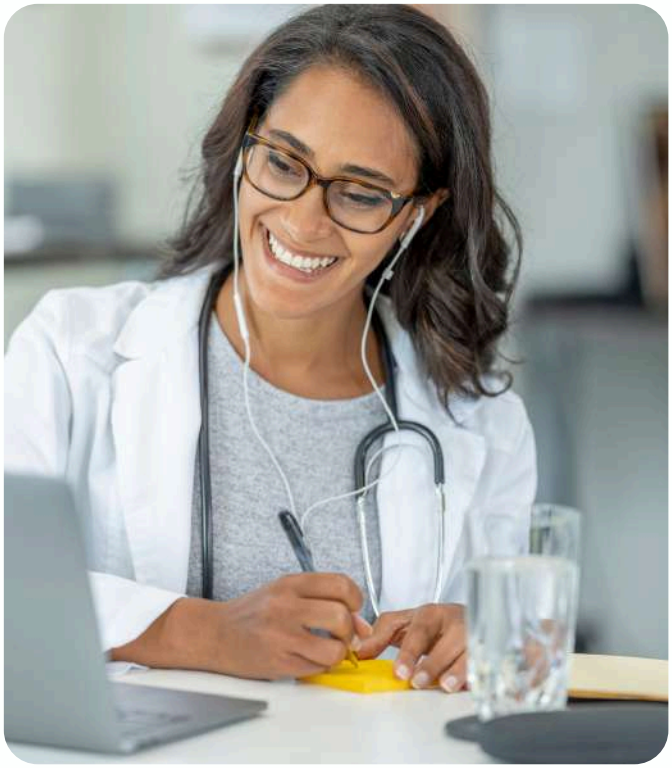
www.Biotecnika.org

Pharmaceutical and Biotechnology Industries

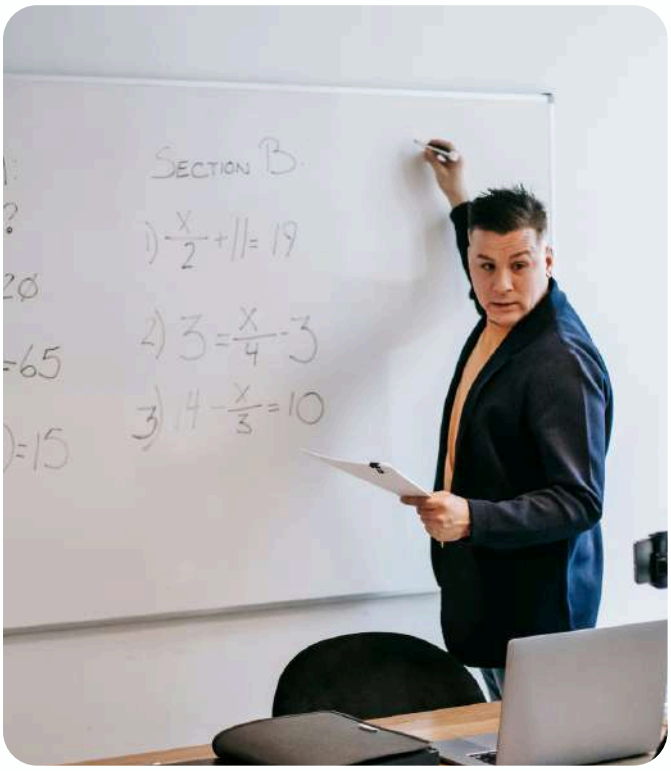
Pharmaceutical and Biotechnology Industries are transforming drug discovery and precision medicine through Data Science. Skilled biologists analyze vast datasets to identify drug targets and improve development. Machine learning models extract insights from molecular interactions, speeding up innovation. This dynamic field offers biologists various career opportunities, establishing them as key contributors to future healthcare.

Healthcare and Clinical Applications

Data Science transforms healthcare by analyzing large patient datasets to uncover disease trends and treatment outcomes. Skilled biologists can develop AI diagnostic tools, predict patient responses, and identify biomarkers for precision medicine. This approach improves understanding of complex diseases and promotes personalized treatment strategies, enhancing patient care and healthcare efficiency.



Academia and Education

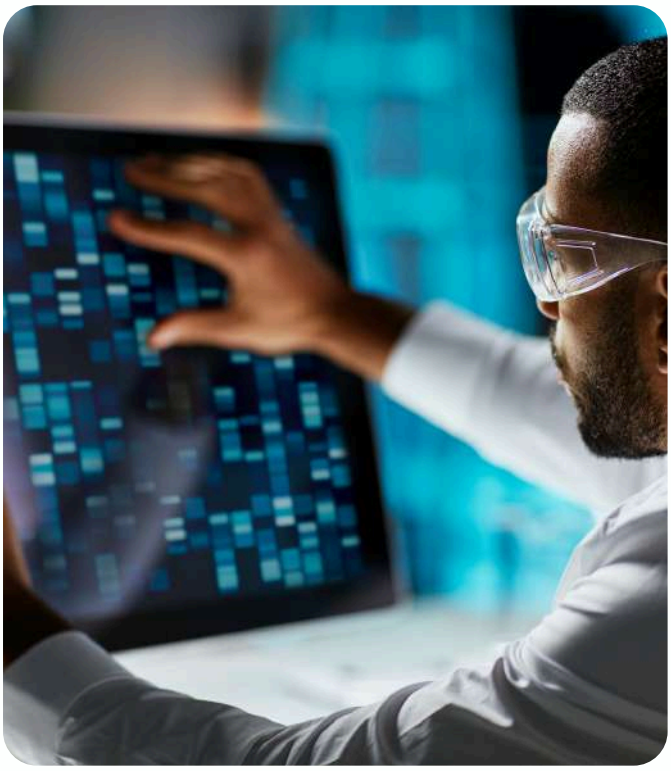


Biologists skilled in Data Science are in high demand in academia, leading interdisciplinary research that combines computational techniques with biology. They enhance genomics, systems biology, and ecological modeling, while also educating future scientists in necessary computational and analytical skills. This expertise fosters impactful research and innovation in the life sciences.

Data Science and Big Data Analytic

Data Science and Big Data Analytics enable researchers to extract insights from large datasets, addressing challenges in biology, genomics, and healthcare. Utilizing machine learning and statistical analysis, data scientists improve agriculture, analyze patient trends, and predict environmental changes, creating diverse career opportunities in academia, industry, and entrepreneurship within the life sciences.

Emerging Technologies



Emerging technologies in data science are transforming biology, creating exciting career opportunities. Biologists are using AI, machine learning, and bioinformatics to revolutionize fields like genomics, drug discovery, agriculture, and healthcare. These innovations enable more precise predictions, enhance research, and drive breakthroughs, opening new paths for biologists to tackle complex challenges and drive innovation.

Global Impact

Data Science is revolutionizing biology globally, enabling breakthroughs in personalized medicine, drug discovery, agriculture, and environmental conservation. By combining computational techniques with biological research, biologists are driving innovation and contributing to global efforts in improving health, sustainability, and well-being.

Data Science provides biologists with numerous career opportunities by leveraging data in genomics, healthcare, agriculture, and environmental science. Their expertise allows biologists to innovate and address complex challenges through data-driven methods. By engaging with this interdisciplinary field, they can enhance their research impact and pursue exciting careers in academia, industry, and entrepreneurship.

How Will This Internship Help You? And What will you learn?

biotecnika
Your Bio Resource

Interns will learn how to use Data Science tools and software to interpret complex biological data, and develop the skills necessary to design and implement their own computational pipelines.

Dive into the refreshing side of science with 30 days, 3 months & 6 months of Data science training! Get a complete 360-degree approach to Data Science with LIVE practicals, assignments & projects

By the end of the program, interns will have gained hands-on experience with cutting-edge techniques and will be well-prepared to pursue a career in Data Science or related fields.

Troubleshooting sessions with seasoned professionals will ensure you have all the support you need to explore this exciting field.

Whether you're just starting out or already have some experience, our Data Science Hands-On Internship is designed to enhance your skills. You'll have the opportunity to learn from industry experts, work with cutting-edge tools and software, and engage in real-world projects. Join us for an exciting journey of learning, professional growth, and hands-on experience!

biotecnika

Your Bio Resource

Become a Data Scientist

Hands-On Internship in Data Science for Biologists

Project | Paper Publication | Work Experience

stores.biotecnika.org | Mail: info@biotecnika.org