**DATA SCIENCE TECHNICAL SKILLS**

**1. Programming Language**

* **Python: Widely used for data manipulation and analysis. Key libraries include:**
  + **Pandas: Data manipulation and analysis.**
  + **NumPy: Numerical computing.**
  + **Scikit-learn: Machine learning algorithms.**
  + **Matplotlib/Seaborn: Data visualization.**
  + **TensorFlow/PyTorch: Deep learning frameworks.**
* **R: Primarily used for statistical analysis and visualization.**
  + **Tidyverse: Data manipulation and visualization.**
  + **ggplot2: Data visualization.**
  + **Caret: Machine learning.**
* **SQL: Essential for database querying and management.**

**2. Machine Learning**

* **Supervised Learning: Algorithms such as linear regression, decision trees, and random forests.**
* **Unsupervised Learning: Clustering (K-means, hierarchical clustering) and dimensionality reduction (PCA).**
* **Deep Learning: Using neural networks for image and text processing.**
  + **Keras: High-level API for building neural networks.**
  + **XGBoost/LightGBM: Gradient boosting algorithms.**

**3. Data Manipulation and Cleaning**

* **Data Wrangling: Techniques for cleaning and preparing data for analysis.**
* **Data Transformation: Using libraries like Dask for parallel data processing.**

**4. Data Visualization**

* **Matplotlib, Seaborn: For creating static and interactive visualizations.**
* **Tableau/Power BI: Business intelligence tools for dashboards.**

**5. Statistical Analysis**

* **Hypothesis Testing: Understanding p-values, confidence intervals, and statistical significance.**
* **Statistical Modelling: Using R and Python for regression analysis and A/B testing.**

**6. Big Data Technologies**

* **Apache Hadoop: Framework for distributed storage and processing of large data sets.**
* **Apache Spark: Unified analytics engine for big data processing.**
* **Hive/Pig: Tools for querying and processing data in Hadoop.**

**7. Natural Language Processing (NLP)**

* **NLTK, SpaCy: Libraries for text processing.**
* **Hugging Face Transformers: Pre-trained models for NLP tasks.**

**8. Cloud Computing**

* **AWS: Services like S3 (storage), EC2 (compute), and SageMaker (machine learning).**
* **Google Cloud Platform: BigQuery, AutoML for scalable data solutions.**
* **Microsoft Azure: Azure Machine Learning and Data Lake.**

**9. Data Engineering Skills**

* **ETL (Extract, Transform, Load): Building data pipelines.**
* **API Integration: Working with RESTful APIs.**

**10. Version Control**

* **Git: For version control and collaboration on coding projects.**

**CERTIFICATION FOR DATA SCIENCE**

**1. IBM Data Science Professional Certificate**

* A comprehensive program covering data analysis, machine learning, and data visualization using Python.

**2. Google Data Analytics Professional Certificate**

* Focuses on data analysis using tools like R and SQL, and includes hands-on projects.

**3. Microsoft Certified: Azure Data Scientist Associate**

* Focuses on implementing machine learning models on the Azure platform.

**4. Certified Data Scientist (DataCamp)**

* A versatile certification covering Python and R, SQL, and machine learning.

**5. TensorFlow Developer Certificate**

* Validates skills in building and training models using TensorFlow.

**6. AWS Certified Machine Learning – Specialty**

* For professionals using AWS to implement machine learning solutions.

**7. SAS Certified Data Scientist**

* Focuses on machine learning, data manipulation, and programming using SAS.

**8. Google Professional Data Engineer**

* Covers designing and managing data processing systems and machine learning models on Google Cloud.

**9. Cloudera Certified Data Scientist (CCP Data Engineer)**

* Focuses on big data tools and data modeling.

**10. Data Science Specialization (Coursera - Johns Hopkins University)**

* Comprehensive program covering data science concepts using R.