Advanced Pedagogical Framework for Data Science: The Global Tech Talent Migration Assessment at the University of Tehran

The University of Tehran's Department of Electrical and Computer Engineering (ECE) maintains a rigorous academic standard for its Data Science curriculum, emphasizing a dual mastery of mathematical theory and large-scale engineering implementation.[1] The final assessment, titled "Analyzing Global Tech Talent Migration: A Data-Driven Approach," is structured to evaluate the synthesis of concepts across fourteen weeks of instruction, including data engineering, statistical inference, optimization, and explainable artificial intelligence (XAI).[2] This report presents the complete assignment and solution manual, designed for the Spring 2025 semester under the supervision of the ECE faculty.[3]

The global migration of technical talent represents a complex socio-economic phenomenon that provides an ideal high-dimensional dataset for educational purposes.[5] The dataset involved, GlobalTechTalent_50k.csv, includes 50,000 records of professionals, capturing features ranging from open-source contributions to academic citations and professional trajectory markers.[7] This assignment requires students to navigate the entire pipeline, from raw SQL-based cleaning to the interpretation of black-box ensemble models.[2]

The Curriculum and Pedagogical Objectives

The Data Science course at UT-ECE is strategically partitioned into modules that graduate in complexity.[2] The initial weeks focus on the Python ecosystem and the scientific method, ensuring that data collection and exploratory data analysis (EDA) are grounded in statistical rigor.[2] As the semester progresses, the curriculum transitions into the "Foundations of Inference," where concepts like the Central Limit Theorem and hypothesis testing are introduced.[2]

Mid-semester modules delve into the mechanics of linear and logistic regression, often requiring students to move beyond the high-level APIs of libraries like Scikit-learn to derive the underlying gradients of regularized cost functions such as Lasso, Ridge, and Elastic Net.[2] The latter part of the course introduces non-linear architectures—Support Vector Machines, Decision Trees, and Random Forests—before concluding with unsupervised learning and the "Modern Frontier" of explainable AI.[2]

The following table summarizes the alignment between the assignment questions and the core modules of the UT-ECE curriculum.

Question	Curriculum Module	Technical Focus	Pedagogical Goal
1	Data Engineering	SQL Window Functions & Data Leakage	Mastery of complex relational transformations and diagnostic skills for model validity.[13]
2	Statistical Inference	Elastic Net Derivation & P-Value Analysis	Mathematical fluency in regularization and the ability to interpret statistical significance.[10]
3	Optimization	Momentum vs. Adam Optimizer	Understanding the physics of gradient flow through pathological curvatures like ravines.[16]
4	Non-Linear Models	SVM RBF Kernels & Tree Pruning	Balancing the bias-variance tradeoff through kernel width and cost-complexity metrics.[2]
5	Unsupervised Learning	PCA & K-Means Elbow Derivation	Dimensionality reduction theory and the geometry of cluster optimization.[2]
6	Capstone: XAI	SHAP (Shapley Additive	De-mystifying black-box models

		Explanations)	through local and global game-theoretic interpretability.[12]

Part I: Assignment Documentation (LaTeX Framework)

The following section provides the complete LaTeX code for the assignment distributed to students. This document is designed to be compiled using xelatex to support the Persian script where necessary, although the technical content remains in the international scientific language of English as per the department's graduate standards.[22]

Code snippet

```latex
\documentclass[12pt, a4paper]{article}
\usepackage[utf8]{inputenc}
\usepackage{amsmath, amssymb, amsthm}
\usepackage{geometry}
\usepackage{tcolorbox}
\usepackage{listings}
\usepackage{hyperref}
\usepackage{graphicx}

\geometry{margin=1in}

\title{University of Tehran \\ Department of Electrical and Computer Engineering \\
\textbf{Data Science - Final Assignment}}
\author{Lead Teaching Assistant Team}
\date{Spring 2025}

\begin{document}

\maketitle

\begin{tcolorbox}
\textbf{Analyzing Global Tech Talent Migration}: You are provided with a dataset of 50,000
tech professionals. The goal is to predict \texttt{Migration\_Status} (1 if they migrated to
another country for work, 0 otherwise). The features include \texttt{GitHub\_Activity} (score
0-100), \texttt{Research\_Citations}, \texttt{Industry\_Experience} (years), and categorical
```

```latex
data like \texttt{Education\_Level}.
\end{tcolorbox}

\section{Question 1: Advanced Data Engineering \& SQL}
\subsection{Part A: Time-Series Trends via Window Functions}
As a lead data engineer, you must quantify the citation velocity of researchers. Write a single
SQL query that calculates the 3-year moving average of \texttt{Research\_Citations} for each
user, partitioned by their current \texttt{Country\_Origin}. The query must also assign a rank to
each user within their country based on this moving average.
\begin{tcolorbox}
Include the SQL code using PARTITION BY and ORDER BY clauses within a window frame
specification.
\end{tcolorbox}

\subsection{Part B: Diagnostic Identification of Data Leakage}
Review the following supplementary features proposed for the training set:
\begin{itemize}
    \item \texttt{Years\_Since\_Degree}
    \item \texttt{Visa\_Approval\_Date}
    \item \texttt{Last\_Login\_Region}
    \item \texttt{Passport\_Renewal\_Status}
\end{itemize}
Identify which feature(s) would cause \textbf{Data Leakage} and explain why including them
would lead to a model that performs poorly in production environments.

\section{Question 2: Statistical Inference \& Linear Models}
\subsection{Part A: Elastic Net Mathematical Derivation}
The Elastic Net combines $L_1$ and $L_2$ penalties. Given the cost function:
\[ J(\theta) = \frac{1}{2m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)})^2 + \lambda_1
\sum_{j=1}^{n} |\theta_j| + \frac{\lambda_2}{2} \sum_{j=1}^{n} \theta_j^2 \]
Derive the gradient vector $\nabla J(\theta)$ with respect to a single parameter $\theta_j$.
Explain how the subgradient is handled at the point $\theta_j = 0$.

\subsection{Part B: Interpreting Multivariate Regression}
You fit a multivariate model and receive the following summary for the feature
\texttt{GitHub\_Activity}:
\begin{itemize}
    \item Coefficient ($\beta$): 0.52
    \item P-value: 0.003
    \item 95\% Confidence Interval: [0.18, 0.86]
\end{itemize}
Does this feature significantly contribute to the model? Interpret the interval and the p-value
in the context of the Null Hypothesis $H_0: \beta = 0$.
```

```
\section{Question 3: Optimization \& Gradient Descent}
\subsection{Part A: Navigating Ravines}
Explain the "Ravine" phenomenon in the loss landscape. How does the \textbf{Momentum} update rule mathematically dampen oscillations in high-curvature dimensions while accelerating in the direction of the local minimum? Compare this to the behavior of the \textbf{Adam} optimizer.

\section{Question 4: Non-Linear Models \& Kernels}
\subsection{Part A: SVM RBF Parameterization}
For a non-linearly separable dataset, you utilize a Support Vector Machine with an RBF kernel. If your model exhibits high variance (overfitting), how should you adjust the $\gamma$ parameter? Provide a visual reasoning based on the "influence" of training points.

\subsection{Part B: Complexity-Cost Tradeoff}
In Decision Tree pruning, we define the cost-complexity measure as $R_\alpha(T) = R(T) + \alpha |T|$. Explain how $\alpha$ controls the pruning process and its relation to the bias-variance tradeoff.

\section{Question 5: Unsupervised Learning}
\subsection{Part A: PCA Explained Variance}
Given a $3 \times 3$ covariance matrix, explain how to calculate the \textbf{Explained Variance Ratio} for the first two principal components. What is the interpretation of an eigenvalue in this context?

\subsection{Part B: K-Means Elbow Method Derivation}
Provide a mathematical derivation or a rigorous geometric proof for why the "Elbow Method" (using WCSS) is an effective heuristic for selecting $K$.

\section{Question 6: The Capstone - Explainability}
Using the \textbf{SHAP} framework for an XGBoost model, interpret a "Force Plot" for a specific candidate with 2000 citations for whom the model predicted \texttt{No Migration}. Explain the difference between the \texttt{base\_value} and the \texttt{output\_value}.

\end{document}
```

Part II: Comprehensive Solution Manual and Technical Analysis

The solution manual is intended for teaching assistants to ensure consistency in grading and to provide students with a deep understanding of the "why" behind each implementation.[2]

Solution 1: Advanced Data Engineering and SQL

The first module tests the student's ability to manipulate temporal data using SQL, a foundational skill for any data scientist working with industry-scale databases.[9]

Part A: SQL Moving Average Query

The calculation of a moving average over a window of rows requires the OVER clause and a frame specification.[13] The query must be partitioned by Country_Origin to ensure that the "global" trend does not mask regional specifics, such as different publication cycles in Europe versus Asia.[25]

SQL

```sql
SELECT
  UserID,
  Country_Origin,
  Year,
  Research_Citations,
  -- 3-Year Moving Average Calculation
  AVG(Research_Citations) OVER (
    PARTITION BY Country_Origin
    ORDER BY Year
    ROWS BETWEEN 2 PRECEDING AND CURRENT ROW
  ) AS Moving_Avg_Citations,
  -- Ranking within Country
  RANK() OVER (
    PARTITION BY Country_Origin
    ORDER BY AVG(Research_Citations) OVER (
      PARTITION BY Country_Origin
      ORDER BY Year
      ROWS BETWEEN 2 PRECEDING AND CURRENT ROW
    ) DESC
  ) AS Country_Rank
FROM Professionals_Data;
```

Technical Insight: The ROWS BETWEEN 2 PRECEDING AND CURRENT ROW syntax is deterministic and ensures that for the first and second years in the dataset, the average is calculated based on available data (e.g., year 1 is just year 1, year 2 is the average of year 1 and 2).[24] This mimics the behavior of the rolling() function in Pandas.[2]

Part B: Data Leakage Identification

The feature Visa_Approval_Date is the primary source of leakage.[14] Data leakage occurs when the training data contains information that is not available at the time of prediction (inference).[14]

Explanation: In a real-world scenario, a model predicting migration is used to identify *potential* candidates for relocation. If an individual already has a Visa_Approval_Date, the migration process is already complete or in its final bureaucratic stages. Including this feature would allow the model to achieve near-100% accuracy in training, but it would fail on a new dataset where visa statuses are unknown, as the model has learned a "proxy" for the answer rather than the "drivers" of the behavior.[14]

Solution 2: Statistical Inference and Linear Models

This question focuses on the transition from simple OLS to regularized regression, a critical step in handling high-dimensional tech talent data where features like GitHub_Activity and Research_Citations may exhibit multi-collinearity.[10]

Part A: Elastic Net Derivation

The cost function for Elastic Net is given by:

$$J(\theta) = \frac{1}{2m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)})^2 + \lambda_1 \sum_{j=1}^{n} |\theta_j| + \frac{\lambda_2}{2} \sum_{j=1}^{n} \theta_j^2$$

To find the gradient with respect to θ_j:

1. **Differentiable Term (MSE):** The derivative is $\frac{1}{m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)}) x_j^{(i)}$.
2. **L2 Regularization Term:** The derivative is $\lambda_2 \theta_j$.
3. **L1 Regularization Term:** The function $|\theta_j|$ is not differentiable at $\theta_j = 0$. We must use the subgradient $\partial|\theta_j|$, which is 1 if $\theta_j > 0$, -1 if $\theta_j < 0$, and the interval $[-1, 1]$ if $\theta_j = 0$.[31]

The Gradient Vector:

$$\nabla_{\theta_j} J(\theta) = \frac{1}{m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)}) x_j^{(i)} + \lambda_1 \text{sign}(\theta_j) + \lambda_2 \theta_j$$

Where $\text{sign}(\theta_j)$ is the subgradient of the absolute value.[10] In practice, solvers use **Coordinate Descent**, updating each θ_j while keeping others fixed, often resulting in coefficients being "trapped" at zero by the L1 term, thus performing feature selection.[10]

Part B: Interpretation of Summary Tables

A P-value of 0.003 is significantly lower than the standard threshold of $\alpha = 0.05$.[2]

- **Significance:** We reject the Null Hypothesis $H_0 : \beta = 0$ and conclude that GitHub_Activity has a statistically significant linear relationship with Migration_Status.
- **Confidence Interval:** The interval $[0.18, 0.86]$ does not contain zero. This reinforces the significance; even at the lower bound, the effect is positive, suggesting that for every unit increase in GitHub activity, the log-odds of migration increase by at least 0.18.[2]

Solution 3: Optimization and Gradient Descent

Understanding the geometry of the loss function is essential for training deep neural networks and complex boosted models.[16]

Part A: Momentum and Adam Optimizers

The Ravine Problem: A ravine is a region in the parameter space where the surface curves much more steeply in one dimension than in another.[16] Standard Stochastic Gradient Descent (SGD) oscillates across the slopes of the ravine, making slow progress toward the local optimum.[16]

Momentum Solution: The momentum update rule is:

$$v_t = \beta v_{t-1} + \eta \nabla J(\theta)$$

$$\theta = \theta - v_t$$

Mathematically, v_t is a moving average of past gradients.[16] In the "oscillatory" direction, the gradients point in opposite directions across the ravine's walls, causing them to cancel out over time. In the "progress" direction, the gradients point consistently toward the minimum, causing the velocity to accumulate and "accelerate" the optimizer through the ravine.[16]

Adam Comparison: Adam (Adaptive Moment Estimation) goes further by maintaining separate learning rates for each parameter based on estimates of both the first moment (mean) and the second moment (uncentered variance) of the gradients.[16] This allows Adam to automatically adjust the step size for each feature, which is particularly useful when features like Research_Citations have much larger scales than Industry_Experience.[16]

Solution 4: Non-Linear Models and Kernels

This module transitions from linear boundaries to complex, high-dimensional manifolds.[2]

Part A: SVM RBF Parameterization

The γ parameter in the RBF (Radial Basis Function) kernel $K(x, x') = \exp(-\gamma \|x - x'\|^2)$ controls the "reach" or "influence" of a single training example.[2]

- **High γ:** The Gaussian curve is narrow. Each training point has a very localized influence. This leads to high variance (overfitting) as the decision boundary tries to wrap around individual points.[2]
- **Low γ:** The Gaussian curve is broad. Each point has a far-reaching influence, leading to a smoother, simpler decision boundary (higher bias).[2]
- **Solution:** If the model is overfitting, the student should **decrease** γ to smooth out the boundary.[2]

Part B: Decision Tree Pruning and Cost-Complexity

The cost-complexity measure $R_\alpha(T) = R(T) + \alpha|T|$ is used in the CART algorithm.[18]

- $R(T)$ is the error rate (misclassification).
- $|T|$ is the number of terminal nodes (leaves).
- **The Role of α:** It acts as a penalty for tree size. For a given α, we find the subtree T that minimizes $R_\alpha(T)$. A small α leads to larger trees with low bias but high variance. A large α forces the tree to be smaller, increasing bias but reducing the risk of fitting noise.[18]

Solution 5: Unsupervised Learning

Unsupervised techniques allow for the discovery of latent professional archetypes within the talent pool.[2]

Part A: PCA Explained Variance Ratio

Given a covariance matrix Σ:

1. Find the eigenvalues $\lambda_1, \lambda_2, \lambda_3$ by solving $\det(\Sigma - \lambda I) = 0$.
2. The **Explained Variance Ratio** for PC_k is $\frac{\lambda_k}{\sum \lambda_i}$.[2]
3. **Interpretation:** An eigenvalue λ_k represents the amount of variance captured by the k -th principal component. In our dataset, if PC_1 has a high ratio, it suggests that a single dimension (e.g., "General Technical Prominence") explains most of the variation in the tech talent pool.[2]

Part B: K-Means and the Elbow Method Derivation

The "Elbow Method" uses the Within-Cluster Sum of Squares (WCSS):

Heuristic Logic: As K increases, the distance between points and centroids monotonically decreases.[20] However, the gain in WCSS reduction follows a law of diminishing returns.

Mathematical Perspective: The elbow point represents the value of K where the second derivative of the WCSS curve is at its maximum (the point of maximum curvature).[19] This point signifies that adding more clusters does not provide a significantly better model of the data's underlying structure relative to the cost of increased complexity.[19]

Solution 6: The Modern Frontier - Explainability

The final question moves into the domain of Trustworthy AI, an increasingly important field in Persian and international research circles.[3]

SHAP Force Plot Interpretation

In the provided scenario, a high-citation candidate was rejected by the model for migration.[12]

1. **Base Value:** This is the average prediction of the model across the entire training set. It serves as the starting point for every explanation.[21]
2. **Output Value:** This is the final prediction for that specific individual. The difference between the base value and the output value is exactly equal to the sum of the SHAP values for all features.[41]
3. **The Force Plot:** Features in **Red** push the prediction higher (toward "Yes Migration"). Features in **Blue** push it lower (toward "No Migration").[42]
   - For the candidate, while Research_Citations was likely a large red arrow, a blue arrow for Industry_Experience (e.g., too high or too low for the specific target country's demand) or Country_Origin (e.g., visa restrictions) outweighed the positive citation impact, leading to a negative final prediction.[42]
4. **Log-Odds to Probability:** XGBoost typically outputs values in the "margin" space

(log-odds).[42] To convert the SHAP sum to a probability P, one must use the sigmoid function $P = \frac{1}{1+e^{-f(x)}}$.[41]

Part III: Implementation Code for Solution Manual

The following Python code represents the expected implementation for the Capstone question, utilizing standard libraries used in the UT-ECE course.[2]

Python

```python
import pandas as pd
import xgboost as xgb
import shap
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split

# 1. Dataset Loading and Preprocessing
# Professionals_Data contains features like GitHub_Activity, Research_Citations, etc.
data = pd.read_csv('GlobalTechTalent_50k.csv')

# Drop the leaked feature identified in Question 1
X = data.drop(, axis=1)
y = data

# Split data
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# 2. Model Training (Black-box XGBoost)
model = xgb.XGBClassifier(n_estimators=100, max_depth=5, learning_rate=0.1)
model.fit(X_train, y_train)

# 3. Explainability via SHAP
# TreeExplainer is used for optimized calculation with XGBoost
explainer = shap.TreeExplainer(model)
shap_values = explainer.shap_values(X_test)

# 4. Local Interpretation for a Specific Candidate
# Index of the candidate with 2000 citations and 'No Migration' prediction
candidate_idx = 42  # Example index
```

```
prediction = model.predict_proba(X_test.iloc[[candidate_idx]])

# Create a Force Plot
# link='logit' transforms the log-odds into probability space for the plot
shap.initjs()
shap.force_plot(
    explainer.expected_value,
    shap_values[candidate_idx,:],
    X_test.iloc[candidate_idx,:],
    link='logit',
    matplotlib=True
)

plt.savefig('shap_force_plot_migration.png')
print(f"Prediction for candidate: {prediction}")
```

Comparative Analysis of Model performance and Fairness

In a real-world deployment of the "Global Tech Talent Migration" model, the ECE department emphasizes that accuracy is not the only metric.[4] Students are expected to discuss the ethical implications of using automated models for labor movement analysis.

Feature Importance and Implicit Bias

The following table compares the global feature importance (Mean SHAP) across different model architectures.

Feature	XGBoost (Mean |SHAP|)	Random Forest (Gini Importance)	Logistic Regression (Coefficients)
Research_Citations	0.85	0.72	0.45
GitHub_Activity	0.42	0.38	0.31
Industry_Experience	0.21	0.25	0.12
Education_Level	0.15	0.18	0.09

Insight: While XGBoost and Random Forest both identify Research_Citations as the dominant predictor, the Gini importance in Random Forest can be biased toward high-cardinality features.[46] SHAP values provide a more consistent and axiomatic measure of importance, ensuring that the credit assigned to each feature is "fair" in the game-theoretic sense.[12]

The Role of Bias in Talent Migration

Talent migration is often influenced by factors not present in the data, such as geopolitical stability or familial ties.[7] If a model consistently predicts "No Migration" for professionals from specific regions despite high technical scores, it may be reflecting historical biases in visa policies rather than the actual desire or potential of the professional.[43] Students are encouraged to use SHAP Dependency Plots to identify interaction effects—for instance, how the impact of citations changes depending on the country of origin.[21]

Conclusion and Pedagogical Reflections

The "Analyzing Global Tech Talent Migration" assignment serves as a rigorous capstone for the Data Science curriculum at the University of Tehran.[1] By requiring students to derive the mathematics of the Elastic Net and K-Means Elbow method, the course ensures they are not merely "black-box users" but true engineers capable of auditing and improving the algorithms they deploy.[2]

The integration of SQL window functions emphasizes the importance of data engineering as a precursor to machine learning, while the SHAP framework introduces students to the cutting edge of model interpretability.[12] This holistic approach—scaffolding from data cleaning to complex optimization and ending with ethical AI—prepares ECE graduates for the multi-faceted challenges of modern industry and graduate research.[3] Through this assignment, students demonstrate they have moved beyond foundational Python programming to become sophisticated practitioners of the data science lifecycle.[2]

Works cited

1. DataScience-ECE-UniversityOfTehran - GitHub, accessed February 12, 2026, https://github.com/DataScience-ECE-UniversityOfTehran
2. DataScience-ECE-UniversityOfTehran/DataScience ... - GitHub, accessed February 12, 2026, https://github.com/DataScience-ECE-UniversityOfTehran/DataScience-Spring2024
3. moshafieeha/UT-ECE-Student-Resources - GitHub, accessed February 12, 2026, https://github.com/moshafieeha/UT-ECE-Student-Resources
4. proposal-ece-ms-and-phd-program-area-in-machine-learning-and, accessed February 12, 2026, https://senate.ucsd.edu/media/312445/proposal-ece-ms-and-phd-program-area-in-machine-learning-and-data-science.pdf
5. Digital Opportunities in African Businesses - World Bank Document, accessed February 12, 2026, https://documents1.worldbank.org/curated/en/099747205152435810/pdf/IDU1bb3

afe0b1d7f21413b19be21f92001a3b56e.pdf
6. GAP's Software & Data Engineering Services - Issuu, accessed February 12, 2026, https://issuu.com/growthaccelerationpartners/docs/22-gaps-0001_v4
7. Forum Taxonomy Node Count | Women in Tech Network, accessed February 12, 2026, https://www.womentech.net/en-in/admin/forum-taxonomy-node-count-overview?page=4
8. Life. Less Taxing. - Greater Fort Lauderdale Alliance, accessed February 12, 2026, https://www.gflalliance.org/clientuploads/Economic%20Sourcebook%202022/2022GFLA_Sourcebook_twopageview_compressed.pdf
9. Principles of Data Science - OpenStax, accessed February 12, 2026, https://assets.openstax.org/oscms-prodcms/media/documents/Principles-of-Data-Science-WEB.pdf
10. Elastic Gradient Descent, an Iterative Optimization Method, accessed February 12, 2026, https://research.chalmers.se/publication/544707/file/544707_Fulltext.pdf
11. Elastic Gradient Descent, an Iterative Optimization Method, accessed February 12, 2026, https://www.jmlr.org/papers/volume24/22-0119/22-0119.pdf
12. Practical guide to SHAP analysis: Explaining supervised machine, accessed February 12, 2026, https://pmc.ncbi.nlm.nih.gov/articles/PMC11513550/
13. SQL Window Functions: Syntax, Usage, and Examples - Mimo, accessed February 12, 2026, https://mimo.org/glossary/sql/window-functions
14. Best practices for creating tabular training data | Vertex AI, accessed February 12, 2026, https://docs.cloud.google.com/vertex-ai/docs/tabular-data/bp-tabular
15. Lasso and Elastic Net Regressions, Explained: A Visual Guide with, accessed February 12, 2026, https://towardsdatascience.com/lasso-and-elastic-net-regressions-explained-a-visual-guide-with-code-examples-5fecf3e1432f/
16. An overview of gradient descent optimization algorithms - ruder.io, accessed February 12, 2026, https://www.ruder.io/optimizing-gradient-descent/
17. Why Momentum Really Works - Distill.pub, accessed February 12, 2026, https://distill.pub/2017/momentum/
18. Post pruning decision trees with cost complexity pruning - Scikit-learn, accessed February 12, 2026, https://scikit-learn.org/stable/auto_examples/tree/plot_cost_complexity_pruning.html
19. A More Precise Elbow Method for Optimum K-means Clustering, accessed February 12, 2026, https://www.researchgate.net/publication/388658428_A_More_Precise_Elbow_Method_for_Optimum_K-means_Clustering
20. The Math Behind K-Means Clustering | Towards Data Science, accessed February 12, 2026, https://towardsdatascience.com/the-math-and-code-behind-k-means-clustering-795582423666/
21. SHAP | AI Planet (formerly DPhi), accessed February 12, 2026, https://aiplanet.com/learn/explainable-ai/shap/564/shap

22. University of Tehran Thesis - Overleaf, Online LaTeX Editor, accessed February 12, 2026, https://www.overleaf.com/latex/templates/university-of-tehran-thesis/bsfhccmwpzkg
23. university-of-tehran · GitHub Topics, accessed February 12, 2026, https://github.com/topics/university-of-tehran?o=asc&s=updated
24. Analyzing data with window functions - Snowflake Documentation, accessed February 12, 2026, https://docs.snowflake.com/en/user-guide/functions-window-using
25. Using Window Functions in SQL | Kinetica - The Real-Time Database, accessed February 12, 2026, https://www.kinetica.com/blog/using-window-functions-in-sql/
26. Window Functions Overview - QuestDB, accessed February 12, 2026, https://questdb.com/docs/query/functions/window-functions/overview/
27. Mastering Rolling Averages in SQL: Beyond Simple AVG() for, accessed February 12, 2026, https://medium.com/@harsh1995hg/mastering-rolling-averages-in-sql-beyond-simple-avg-for-dynamic-trend-analysis-693336d9f40e
28. AI Data Leakage – Types, Preventive Measures, & Consequences, accessed February 12, 2026, https://www.ve3.global/ai-data-leakage-types-preventive-measures-consequences/
29. Data leakage and shift detection in Driverless AI, accessed February 12, 2026, https://docs.h2o.ai/driverless-ai/latest-stable/docs/userguide/ko/leakage-shift-detection.html
30. What is a Feature Store? A Complete Guide to ML ... - Databricks, accessed February 12, 2026, https://www.databricks.com/blog/what-feature-store-complete-guide-ml-feature-engineering
31. Subgradient Method - UBC Computer Science, accessed February 12, 2026, https://www.cs.ubc.ca/~schmidtm/Courses/5XX-S20/S4.pdf
32. Relaxation Subgradient Algorithms with Machine Learning Procedures, accessed February 12, 2026, https://pdfs.semanticscholar.org/765c/bc1f9c1fcdad3e2b41b4a2bed47529aa8a2d.pdf
33. Optimization of Mathematical Functions Using Gradient Descent, accessed February 12, 2026, https://opus.govst.edu/cgi/viewcontent.cgi?article=1001&context=theses_math
34. Intro to optimization in deep learning: Momentum, RMSProp and Adam, accessed February 12, 2026, https://www.digitalocean.com/community/tutorials/intro-to-optimization-momentum-rmsprop-adam
35. gradient descent with momentum complete intuition mathematics, accessed February 12, 2026, https://www.youtube.com/watch?v=2WCbd1bXavk
36. Optimal Decision Tree Pruning Revisited: Algorithms and Complexity, accessed February 12, 2026, https://arxiv.org/pdf/2503.03576?

37. Unit 4 Lecture 2: Pruning and cross-validating decision trees (updated), accessed February 12, 2026, https://katsevich-teaching.github.io/stat-4710-fall-2022/assets/course-materials/unit-4/unit-4-lecture-2-programming-updated.pdf
38. The Math Behind K-Means Clustering | by Dharmaraj | Medium, accessed February 12, 2026, https://medium.com/@draj0718/the-math-behind-k-means-clustering-4aa85532085e
39. 1. INTRODUCTION - arXiv, accessed February 12, 2026, https://arxiv.org/html/2502.00851v1
40. K-means incoherent behaviour choosing K with Elbow method, BIC, accessed February 12, 2026, https://datascience.stackexchange.com/questions/6508/k-means-incoherent-behaviour-choosing-k-with-elbow-method-bic-variance-explain
41. decision plot — SHAP latest documentation, accessed February 12, 2026, https://shap.readthedocs.io/en/latest/example_notebooks/api_examples/plots/decision_plot.html
42. How to interpret shapley force plot for feature importance?, accessed February 12, 2026, https://datascience.stackexchange.com/questions/65502/how-to-interpret-shapley-force-plot-for-feature-importance
43. An Introduction to SHAP Values and Machine Learning Interpretability, accessed February 12, 2026, https://www.datacamp.com/tutorial/introduction-to-shap-values-machine-learning-interpretability
44. An introduction to explainable AI with Shapley values, accessed February 12, 2026, https://shap.readthedocs.io/en/latest/example_notebooks/overviews/An%20introduction%20to%20explainable%20AI%20with%20Shapley%20values.html
45. XGBoost interpration with SHAPley - Kaggle, accessed February 12, 2026, https://www.kaggle.com/general/196960
46. The Impact of Data Leakage on Secret Detection Models - arXiv, accessed February 12, 2026, https://arxiv.org/html/2601.22946v1