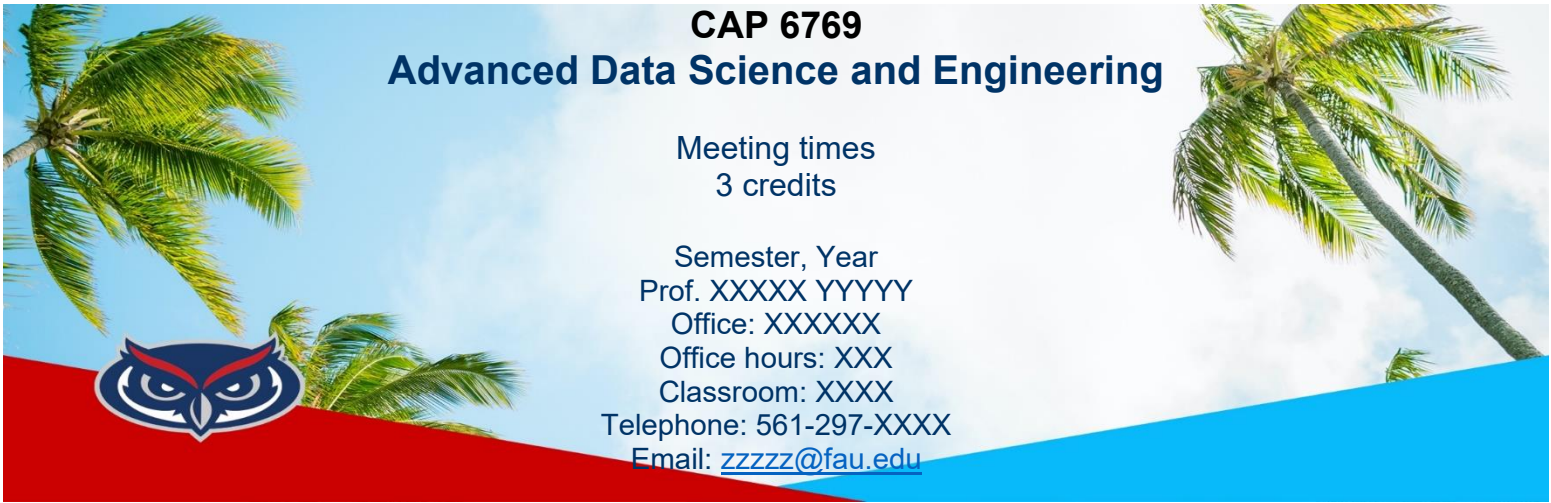
 <b>FLORIDA ATLANTIC UNIVERSITY</b>	<b>NEW COURSE PROPOSAL</b> <b>Graduate Programs</b>		UGPC Approval _____ UFS Approval _____ SCNS Submittal _____ Confirmed _____ Banner _____ Catalog _____	
	<b>Department</b>  <b>College</b> (To obtain a course number, contact <a href="mailto:erudolph@fau.edu">erudolph@fau.edu</a> )			
<b>Prefix</b> CAP  <b>Number</b> 6769	(L = Lab Course; C = Combined Lecture/Lab; add if appropriate) <b>Lab Code</b>	<b>Type of Course</b> Lecture	<b>Course Title</b>	
<b>Credits</b> (See <a href="#">Definition of a Credit Hour</a> )  3	<b>Grading</b> (Select One Option)  <b>Regular</b> <input type="radio"/> <b>Sat/UnSat</b> <input type="radio"/>	<b>Course Description</b> (Syllabus must be attached; see <a href="#">Template</a> and <a href="#">Guidelines</a> ) This advanced course builds on foundational data science and engineering skills to develop expertise in managing, processing, and analyzing complex, large-scale, and real-world datasets. Students will work through the entire data science pipeline, progressing from data import and storage to wrangling, quality assessment, feature engineering, modeling, and communication. Emphasis is placed on reproducibility, ethical data practices, and producing professional stakeholder-ready deliverables. Through hands-on exercises and an end-to-end applied project, students will gain practical experience in designing and implementing robust data workflows for a variety of domains.		
<b>Effective Date</b> (TERM & YEAR)  Spring 2026				
<b>Prerequisites</b>      <i>Prerequisites, Corequisites and Registration Controls are enforced for all sections of course.</i>		<b>Academic Service Learning (ASL) course</b> <input type="checkbox"/> Academic Service Learning statement must be indicated in syllabus and approval attached to this form.		
		<b>Corequisites</b>	<b>Registration Controls</b> (For example, Major, College, Level)	
<b>Minimum qualifications needed to teach course:</b>  Terminal degree in related field.		<b>List textbook information in syllabus or here</b>		
<b>Faculty Contact/Email/Phone</b> Raquel Assis / rassis@fau.edu / 561-297-3927		<b>List/Attach comments from departments affected by new course</b>		

<b>Approved by</b> Department Chair <u>Haik Kalva</u> College Curriculum Chair <u>A.R. Nayak</u> College Dean <u>Raquel Assis</u> UGPC Chair <u>[Signature]</u> UGC Chair <u>[Signature]</u> Graduate College Dean <u>[Signature]</u> UFS President _____ Provost _____	<b>Date</b> 9/30/2025 10/07/2025 10/8/2025 11/05/2025 11/05/2025 11/05/2025 _____ _____
---	---

Email this form and syllabus to [UGPC@fau.edu](mailto:UGPC@fau.edu) 10 days before the UGPC meeting.



# CAP 6769 Advanced Data Science and Engineering

Meeting times  
3 credits

Semester, Year  
Prof. XXXXX YYYYY  
Office: XXXXX  
Office hours: XXX  
Classroom: XXXX  
Telephone: 561-297-XXXX  
Email: [zzzzz@fau.edu](mailto:zzzzz@fau.edu)



TA name	xxxxxx xxxxxxxxxx
Office	xxxxxx
Office hours	MWF xx:xx – xx:xx
Telephone	561-297-xxxx
Email	xxxxxx@fau.edu

## Course Description

This advanced course builds on foundational data science and engineering skills to develop expertise in managing, processing, and analyzing complex, large-scale, and real-world datasets. Students will work through the entire data science pipeline, progressing from data import and storage to wrangling, quality assessment, feature engineering, modeling, and communication. Topics include API generation, non-relational databases, distributed computing frameworks, advanced imputation and resampling methods, high-dimensional visualization, interpretable modeling, and hyperparameter optimization. Emphasis is placed on reproducibility, ethical data practices, and producing professional stakeholder-ready deliverables. Through hands-on exercises and an end-to-end applied project, students will gain practical experience in designing and implementing robust data workflows for a variety of domains.

## Instructional Method

This class is designated as “In-Person w/Recorded Lecture” (section XXX) or “Videotaped Class” (section YYY). In-person class sessions will be automatically recorded and uploaded to Canvas within 24 hours. Student enrolled in section XXX may choose to attend in-person classes or view recordings, whereas students enrolled in section YYY are only able to view recordings.

## Prerequisites/Corequisites

CAP 4773 or 5768

## Course Objectives/Student Learning Outcomes

Upon successful completion of this course, students will be able to:

1. Import, store, and manage complex datasets from heterogeneous sources and formats.

2. Clean, transform, and prepare large-scale data for analysis using advanced wrangling, quality assessment, and feature engineering techniques.
3. Build, optimize, and interpret machine learning models with attention to transparency, bias mitigation, and ethical considerations.
4. Design reproducible, automated data science workflows and pipelines.
5. Communicate analytical processes and results effectively through a professional written report.

## Course Evaluation Method

	Weight
Five homework assignments (10% each)	50%
Project proposal	10%
Final project report	40%
<b>Total</b>	<b>100%</b>

## Course Grading Scale

Grade	Total (%)
A	[93 – 100]
A-	[90 – 93)
B+	[87 – 90)
B	[83 – 87)
B-	[80 – 83)
C+	[77 – 80)
C	[73 – 77)
C-	[70 – 73)
D+	[67 – 70)
D	[63 – 67)
D-	[60 – 63)
F	[0 – 60)

Note: Brackets are inclusive, parentheses are exclusive

## Policy on Makeup Tests, Late Work, and Incompletes (if applicable)

There will be no exams, and therefore no makeup exams, in this course.

Late work will not be accepted. All assignments will be posted well in advance, and students may submit assignments early. Assignments may also be submitted multiple times prior to the due date. Any assignment not turned in by the due date will result in a zero.

Incomplete grades are against the policy of the department, and they will only be assigned if there is solid evidence of a medical or other serious emergency situation and the student is currently passing the class.

## **Policy on the Recording of Lectures**

Students enrolled in this course may record video or audio of class lectures for their own personal educational use. A class lecture is defined as a formal or methodical oral presentation as part of a university course intended to present information or teach students about a particular subject. Recording class activities other than class lectures, including but not limited to student presentations (whether individually or as part of a group), class discussion (except when incidental to and incorporated within a class lecture), labs, clinical presentations such as patient history, academic exercises involving student participation, test or examination administrations, field trips, and private conversations between students in the class or between a student and the lecturer, is prohibited. Recordings may not be used as a substitute for class participation or class attendance and may not be published or shared without the written consent of the faculty member. Failure to adhere to these requirements may constitute a violation of the University's Student Code of Conduct and/or the Code of Academic Integrity.

## **Attendance Policy**

Students are expected to attend all of their scheduled University classes and to satisfy all academic objectives as outlined by the instructor. The effect of absences upon grades is determined by the instructor, and the University reserves the right to deal at any time with individual cases of non-attendance. Students are responsible for arranging to make up work missed because of legitimate class absence, such as illness, family emergencies, military obligation, court-imposed legal obligations or participation in University-approved activities. Examples of University-approved reasons for absences include participating on an athletic or scholastic team, musical and theatrical performances and debate activities. It is the student's responsibility to give the instructor notice prior to any anticipated absences and within a reasonable amount of time after an unanticipated absence, ordinarily by the next scheduled class meeting. Instructors must allow each student who is absent for a University-approved reason the opportunity to make up work missed without any reduction in the student's final course grade as a direct result of such absence.

## **Counseling and Psychological Services (CAPS) Center**

Life as a university student can be challenging physically, mentally and emotionally. Students who find stress negatively affecting their ability to achieve academic or personal goals may wish to consider utilizing FAU's Counseling and Psychological Services (CAPS) Center. CAPS provides FAU students a range of services – individual counseling, support meetings, and psychiatric services, to name a few – offered to help improve and maintain emotional well-being. For more information, go to <http://www.fau.edu/counseling/>

## **Disability Policy**

In compliance with the Americans with Disabilities Act Amendments Act (ADAAA), students who require reasonable accommodations due to a disability to properly execute coursework must register with Student Accessibility Services (SAS) and follow all SAS procedures. SAS has offices across three of FAU's campuses – Boca Raton, Davie and Jupiter – however disability services

are available for students on all campuses. For more information, please visit the SAS website at [www.fau.edu/sas/](http://www.fau.edu/sas/).

## Code of Academic Integrity

Students at Florida Atlantic University are expected to maintain the highest ethical standards. Academic dishonesty is considered a serious breach of these ethical standards, because it interferes with the university mission to provide a high quality education in which no student enjoys an unfair advantage over any other. Academic dishonesty is also destructive of the university community, which is grounded in a system of mutual trust and places high value on personal integrity and individual responsibility. Harsh penalties are associated with academic dishonesty. For more information, see [University Regulation 4.001](#).

## Required Texts/Readings

This course will utilize free online resources and publications from the primary literature. A core reading list is provided below, though other sources may be added as they become available.

Dean J, Ghemawat S. MapReduce: Simplified data processing on large clusters. *Comm ACM* 52:1. 2004.

Sculley D, Holt G, Golovin D, Davydov E, Phillips T. Hidden technical debt in machine learning systems. *Adv Neur Inf Proc Syst* 28. 2015.

Halevy A, Norvig P, Pereira F. The unreasonable effectiveness of data. *IEEE Intel Syst* 24:2. 2009.

Zou H, Hastie T. Regularization and variable selection via the elastic net. *J Royal Stat Soc B* 58: 267-288. 1996.

Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: Synthetic Minority Over-sampling Technique. *arXiv*:1106. 2011.

McInnes L, Healy J, Melville J. UMAP: Uniform Manifold Approximation and Projection for dimension reduction. *arXiv*:1802.03246. 2018.

Lundberg S, Lee S-I. A unified approach to interpreting model prediction. *Adv Neur Inf Proc Syst* 30. 2017.

## Course Topical Outline

### Module 1: Data import and parsing

- Working with APIs
- Importing from non-relational databases
- Importing from streaming sources
- Reading and parsing semi-structured data

## Module 2: Data storage and management

- Data warehousing concepts
- Managing large datasets locally and in the cloud
- Data versioning and reproducibility
- Trade-offs in storage formats and retrieval speed

## Module 3: Data wrangling

- Handling high-dimensional and sparse data
- Complex joins, merges, and reshaping for multi-table datasets
- Cleaning inconsistent identifiers and hierarchical data
- Efficient wrangling with parallel and distributed tools

## Module 4: Data quality and missing data

- Missing data and their implications
- Advanced imputation techniques
- Outlier detection and treatment in large datasets
- Data validation frameworks

## Module 5: Feature engineering at scale

- Encoding high-cardinality categorical variables
- Interaction and polynomial feature generation
- Automated feature selection methods
- Feature engineering from semi-structured fields

## Module 6: Data balancing and resampling

- Class imbalance detection and diagnostics
- Undersampling and oversampling strategies
- SMOTE and advanced synthetic data generation methods
- Evaluating impact of resampling on model generalizability

## Module 7: Visualization and exploratory data analysis

- Visualizing high-dimensional data
- Dynamic and interactive dashboards
- Efficiently visualizing large datasets
- Designing EDA workflows for messy real-world data

## Module 8: Modeling with interpretability

- Interpretability methods
- Selecting interpretable models for stakeholders
- Bias detection and mitigation

## Module 9: Model optimization

- Hyperparameter tuning
- Pipeline automation for modeling
- Preventing overfitting with complex data

## Module 10: Communication, ethics, and reproducibility

- End-to-end workflow documentation and reproducibility best practices
- Ethics and data privacy
- Creating compelling reports for technical and non-technical audiences



Outlook

---

**Re: Support for graduate course proposal**

---

**From** Yuan Wang <YWANG@fau.edu>  
**Date** Thu 9/4/2025 9:17 AM  
**To** Raquel Assis <rassis@fau.edu>

Dear Raquel,

Thank you for reaching out. We support your proposal for the new course, Advanced Data Science and Engineering.

Best regards,  
Yuan

On 8/28/2025 11:59 AM, Raquel Assis wrote:

Dear Yuan,

I hope your semester is starting off smoothly.

I am reaching out to ask if you would be willing to support our attached graduate course proposal for Advanced Data Science and Engineering.

This proposal was developed in response to student interest, as many who have completed CAP 5768 have expressed a strong desire for a more advanced course.

Please let me know if you have any questions or want to discuss this further.

Best,  
Raquel

Raquel Assis, Ph.D.  
Associate Dean for Graduate Studies, Associate Professor, and I-Health Fellow  
College of Engineering and Computer Science  
Florida Atlantic University  
<http://assisgroup.fau.edu/>





---

**Re: Support for graduate course proposal**

---

**From** Tamara Dinev <tdinev@fau.edu>

**Date** Tue 10/7/2025 2:34 PM

**To** Raquel Assis <rassis@fau.edu>

Hi Raquel

I have no objections to the new proposed course.

Thank you!

Tamara

Best Regards:

Tamara Dinev, Ph.D.

Department Chair and Professor

Dean's Distinguished Research Fellow

Department of Information Technology and Operations Management, FL 219

College of Business, Florida Atlantic University

Boca Raton, Florida 33431

Google Scholar: <https://scholar.google.com/citations?user=YH8QZ-YAAAAJ&hl=en>

On Oct 7, 2025 12:00 PM, Raquel Assis <rassis@fau.edu> wrote:

Dear Tamara,

Hope you are doing well. I'm reaching out to ask if you would be willing to support our attached graduate course proposal for Advanced Data Science and Engineering.

This proposal was developed in response to student interest, as many who have completed CAP 5768 have expressed a strong desire for a more advanced course.

Please let me know if you have any questions or want to discuss this further.

Best,  
Raquel

Raquel Assis, Ph.D.

Associate Dean for Graduate Studies, Associate Professor, and I-Health Fellow

College of Engineering and Computer Science

Florida Atlantic University

<http://assisgroup.fau.edu/>










# ncp-cap6769-advanced-data-science-engineerin g

Final Audit Report

2025-11-05

Created:	2025-11-05
By:	Robert Stackman (rstackma@fau.edu)
Status:	Signed
Transaction ID:	CBJCHBCAABAA_4CVxiqjuN7GLTSAIzJcuYB8QGLbY9b

## "ncp-cap6769-advanced-data-science-engineering" History

-  Document created by Robert Stackman (rstackma@fau.edu)  
2025-11-05 - 5:56:24 PM GMT
-  Document emailed to sementel@fau.edu for signature  
2025-11-05 - 5:57:16 PM GMT
-  Email viewed by sementel@fau.edu  
2025-11-05 - 7:41:57 PM GMT
-  Signer sementel@fau.edu entered name at signing as Arthur Sementelli  
2025-11-05 - 7:43:28 PM GMT
-  Document e-signed by Arthur Sementelli (sementel@fau.edu)  
Signature Date: 2025-11-05 - 7:43:30 PM GMT - Time Source: server
-  Document emailed to Robert Stackman (rstackma@fau.edu) for signature  
2025-11-05 - 7:43:34 PM GMT
-  Email viewed by Robert Stackman (rstackma@fau.edu)  
2025-11-05 - 8:39:57 PM GMT
-  Document e-signed by Robert Stackman (rstackma@fau.edu)  
Signature Date: 2025-11-05 - 8:40:11 PM GMT - Time Source: server
-  Agreement completed.  
2025-11-05 - 8:40:11 PM GMT