Subject Description Form

Subject Code	COMP5575
Subject Title	Advanced Techniques for High-Dimensional Data Management and Analytics
Credit Value	3
Level	5
Pre-requisite/ Co-requisite/ Exclusion	Some knowledge in databases and analytics is preferable
Objectives	After successfully completing this course, you should be able to: 1. identify the key characteristics of various complex high-dimensional data types, including time-series, geospatial,spatiotemporal, and multimedia data, compared to traditional relational database; 2. understand and design efficient data management solutions for spatial and spatiotemporal data using state-of-the-artalgorithms and technologies; 3. analyze and apply forecasting techniques for time-series data to predict future trends and patterns in real-world contexts; 4. incorporate machine learning and AI techniques into analysis and applications of multimedia data to derive actionable insights; and 5. gain proficiency in advanced techniques for high-dimensional data, utilizing database management, machine learning, and AI approaches.
Intended Learning Outcomes	Upon completion of the subject, students will be able to: a. describe and classify various complex high-dimensional data types and demonstrate their practical use in real-world scenarios; b. design and implement efficient indexing and querying algorithms for geospatial and spatiotemporal data and validate their efficiency, strengths, and limitations through practical assignments; c. apply and evaluate forecasting techniques for different types of time-series data, such as stock prices, retail sales, and traffic flow, deriving insights through hands-on case studies; d. utilize machine learning and AI techniques to analyze multimedia data, such as image, audio and video data, presenting findings through demonstrations and reports e. create and interpret data visualizations using advanced tools, effectively communicating high-dimensional data insights to

	diverse audiences through visual storytelling and interactive dashboards.
Subject Synopsis/ Indicative Syllabus	• **Introduction to High-Dimensional Data:** overview of high-dimensional data types; comparison with traditional relational databases; real-world application and challenges. • **Geospatial & Spatiotemporal Data:** characteristics and organization of geospatial and spatiotemporal data; diverse trajectory similarity measures; efficient indexing structures and querying algorithms; advanced routing algorithms on road networks. *Assessment: practical assignments on designing and implementing data management solutions.* • **Times-series Data:** introduction to different types of time-series data and their characteristics; time-series data modeling such as recurrent neural networks; forecasting techniques and prediction models. *Assessment: case studies involving real-world data from various industries such as finance, retail, and smart city.* • **Multimedia Data:** introduction to different types of multimedia data and feature extraction; similarity search and information retrieval; integration of machine learning and AI techniques. *Assessment: collaborative projects and demonstrations for multimedia data analytics from social media, medical images, etc.*
Teaching/Learning Methodology	Class activities including lecture, tutorial, lab, workshop seminar where applicable. • **Lectures** will introduce abstract concepts, complex algorithms, and practical applications to help students thoroughly understand and master these high-dimensional data and related knowledge. The focus will be on conceptual and algorithmic understanding. *It will primarily support the intended learning outcomes a) and b).* • **Tutorials** will provide opportunities to practice the knowledge delivered by lectures through short-answer questions and other formats. Students will actively express their understanding and solidify their learning. *It will primarily support the intended learning outcomes of b) and c).* • **Labs** will be tailored to either an engineering-oriented or research-oriented approach based on student preference. Students can choose to implement high-dimensional solutions systematically or design new algorithms or models with academic novelty. *It will primarily support the intended learning outcomes of c), d), and e).* • **Workshop seminar**, if applicable, will invite experts from both academia and industry who will share state-of-the-art ideas and knowledge in high-dimensional data management and analytics in real life. *It will expose students to real-world applications and advancements, supporting all learning outcomes indirectly.*

Assessment Methods in Alignment with Intended Learning Outcomes	Specific assessment methods/tasks	% weighting	Intended subject learning outcomes to be assessed (Please tick as appropriate)				
			a	b	c	d	e
	Assignments and Project	35%	✔	✔	✔	✔	✔
	Mid-term Test	20%	✔	✔	✔		
	Final Examination	45%	✔	✔	✔		
	Total	100 %					

- **Assignments and projects** typically offer a hands-on approach to learning, allowing students to apply theoretical knowledge to practical scenarios. By working on various tasks that align with all five learning outcomes, students can demonstrate their ability to describe, classify, design, implement, apply, evaluate, and visualize high-dimensional data. This continuous assessment method encourages consistent engagement and deeper understanding of the subject matter.

- **The mid-term test** serves as a checkpoint for students to assess their understanding of the initial topics covered, particularly focusing on the classification of data types, algorithm design, and forecasting techniques. This continuous assessment ensures that students have a solid foundation before progressing to more advanced topics.

- **The final exam** evaluates students' comprehensive knowledge and understanding of the entire course content. It ensures that students can synthesize and apply their learning to complex problems. This method is appropriate for assessing students' retention and overall grasp of the subject.

Student Study Effort Expected	Class contact:	
	▪ Class activities (lecture, tutorial, lab)	39 Hrs.
	Other student study effort	
	▪ Assignments, Quizzes, Projects, Exams	66 Hrs.
	Total student study effort	105 Hrs.

Reading List and References	*Textbooks:*
	(1) Mamoulis, Nikos. Spatial data management. Springer Nature, 2022.
	(2) Zheng, Yu, and Xiaofang Zhou. Computing with spatial trajectories. Springer Science & Business Media, 2011.
	(3) Bühlmann, Peter, and Sara Van De Geer. Statistics for high-dimensional data: methods, theory and applications. Springer Science & Business Media, 2011.
	(4) Shmueli, Galit, and Julia Polak. Practical time series forecasting

with r: A hands-on guide. Axelrod schnall publishers, 2024.

(5) Bhatti, Uzair Aslam, et al., eds. Deep Learning for Multimedia Processing Applications: Volume Two: Signal Processing and Pattern Recognition. CRC Press, 2024.

Other references:

(1) Li, Zijian, Lei Chen, and Yue Wang. "G*-tree: An efficient spatial index on road networks." 2019 IEEE 35th International Conference on Data Engineering (ICDE). IEEE, 2019.

(2) Tao, Yaguang, et al. "A comparative analysis of trajectory similarity measures." GIScience & Remote Sensing 58.5 (2021): 643-669.

(3) Shang, Shuo, et al. "Introduction to Spatio-temporal data management and analytics for Smart City research." GeoInformatica 24 (2020): 1-2.

(4) Mahalakshmi, Ganapathy, S. Sridevi, and Shyamsundar Rajaram. "A survey on forecasting of time series data." 2016 international conference on computing technologies and intelligent data engineering (ICCTIDE'16). IEEE, 2016.

(5) Ahmed, Dozdar Mahdi, Masoud Muhammed Hassan, and Ramadhan J. Mstafa. "A review on deep sequential models for forecasting time series data." Applied Computational Intelligence and Soft Computing 2022.1 (2022): 6596397.

(6) Lahat, Dana, Tülay Adali, and Christian Jutten. "Multimodal data fusion: an overview of methods, challenges, and prospects." Proceedings of the IEEE 103.9 (2015): 1449-

(7) Alpaydin, Ethem. "Classifying multimodal data." The Handbook of Multimodal-Multisensor Interfaces: Signal Processing, Architectures, and Detection of Emotion and Cognition-Volume 2. 2018. 49-69.

(8) Tang, Jian, et al. "Visualizing large-scale and high-dimensional data." Proceedings of the 25th international conference on world wide web. 2016.