HEI4S3-RM
Applied Big Data and Machine Learning

Presenter: Antonios Koufou

Organisation: International Hellenic University

HEI4S3-RM

Building Ecosystem Integration
Labs at HEI to foster
Smart Specialization
and Innovation
on Sustainable
Raw Materials

Overview

Introduction - general

- A few words about Big Data

- General about Machine Learning - Applications

Case Studies

- Big Data in Occupational Safety

- Big Data in Computational Chemistry

- Machine Learning in Occupational Safety

- Machine Learning in Computational Chemistry

Introduction - General

- Sections will include key topics that will be related to questions in the assessment

- Assessment will take place after the end of the session in Google forms and will consist of combined questions for all three modules (13+14 February sessions)

- Question important topics will be marked with a discrete RED and UNDERLINED text throughout slides

- Viewers should be aware to take notice of these sample important topics

1 Big Data – few words

Big Data field of action

- Big Data and Analytics is thought to be about:

- Business Intelligence and Analytics

- Computational Science

- But, it is much more than that!

- Demographic Analysis

- Geointelligence: Spatial Analysis

- The Grand Challenges in Science

- Bioscience: Processing 3-D hyperspectral high resolution images for diagnostics, genomic research, Proteonomics, etc.

- Media Analysis: Processing text, audio, video, imagery

- equipment and environment

- And much, much more …..

- However…. <u>Big Data is NOT focused on technology alone</u>

1 Big Data – few words

- No standard definition!

- Wikipedia: Big data is a term for data sets that are so large or complex that traditional data processing applications are inadequate.

- Challenges include analysis, capture, data curation, search, sharing, storage, transfer, visualization, querying, updating and information privacy.

- Analysis of data sets can find new correlations to "spot business trends, prevent diseases, combat crime and many ... many more."
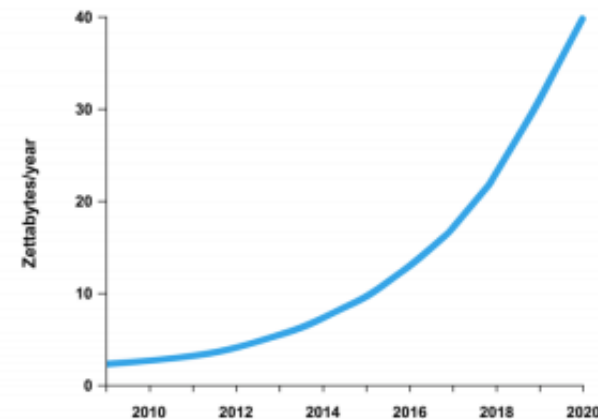
1 Big Data – few words

- Who's generating Big Data?

- Social Media and Networks

- Scientific Research

- Mobile Devices

- Sensor Technology and Networks

- The progress and innovation is no longer hindered by the ability to collect data

- But, by the ability to manage, analyze, summarize, visualize, and discover knowledge from the collected data in a timely manner and in a scalable fashion

1 Big Data – few words

- Huge amount of Data created worldwide
- Data Volume
- Growth 40% per year
- From 8 zettabytes (2016) to 44zb (2020)

- Data volume is increasing exponentially

Size of the Digital Universe – Annual Data Created & Consumed

Data Source: IDC Digital Universe 2013

1 Big Data – few words

- Examples

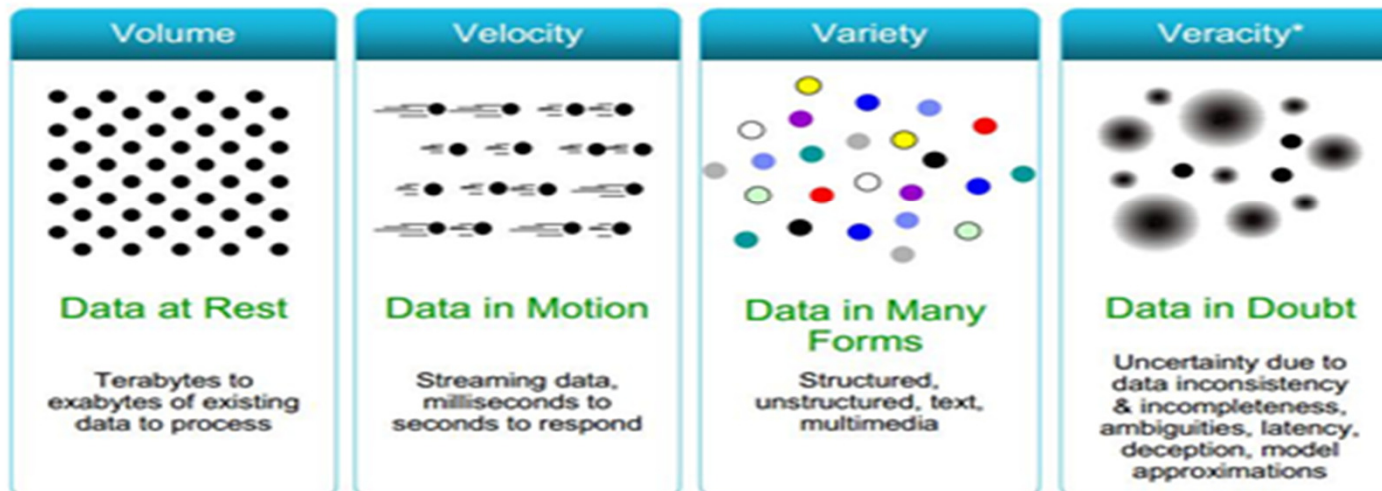- Internet Archive: more than 10PB

- CERN LHC: 15PB/year

1 Big Data – 3(?) Vs

- Big data is usually transformed in three dimensions- volume, velocity and variety.

- Volume: Machine generated data is produced in larger quantities than non traditional data.

- Velocity: This refers to the speed of data processing.

- Variety: This refers to large variety of input data which in turn generates large amount of data as output.

1 Big Data – 3(?) Vs – or more

Some make it 4 v's

Volume	Velocity	Variety	Veracity*
Data at Rest	**Data in Motion**	**Data in Many Forms**	**Data in Doubt**
Terabytes to exabytes of existing data to process	Streaming data, milliseconds to seconds to respond	Structured, unstructured, text, multimedia	Uncertainty due to data inconsistency & incompleteness, ambiguities, latency, deception, model approximations

1 Big Data – or … more Vs

- The extra 3 Vs

- Veracity: Data quality issues are particularly challenging in a big data context.

- Visibility/Visualization: After big data has been processed, we need a way of presenting the data in a manner that's readable and accessible.

- Value: Ultimately, big data is meaningless if it does not provide value toward some meaningful goal.

1 Big Data – Velocity (speed)

■ Data is generated fast and needs to be processed fast

■ Online Data Analytics

■ Late decisions equals missing opportunities

■ Examples

- E-Promotions: Based on your current location, your purchase history, what you like send promotions right now for store next to you

- Healthcare monitoring: sensors monitoring your activities and body - any abnormal measurements require immediate reaction

- Disaster management and response

1 Big Data Tools (general)

- Microsoft Azure HDInsight: Azure HDInsight is a fully managed cloud Big Data platform offered by Microsoft Azure. It provides Apache Hadoop, Apache Spark, Apache HBase, Apache Storm, and other Big Data technologies as managed services on the Azure cloud platform.

- Microsoft SQL Server Big Data Clusters: SQL Server Big Data Clusters is an integrated Big Data solution provided by Microsoft SQL Server.

- IBM BigInsights for Apache Hadoop: IBM BigInsights is an enterprise-grade Hadoop distribution provided by IBM.

- Cloudera Data Platform Private Cloud: Cloudera Data Platform is a comprehensive Big Data platform provided by Cloudera.

- MapR Data Platform: MapR Data Platform is a converged data platform that combines Hadoop, Spark, and other Big Data technologies for real-time data processing, analytics, and machine learning.

2

Machine Learning – a few words

Why "LEARNING"?

- Machine learning is about programming computers to optimize a performance criterion using example data or past experience.

- There is no need to "learn" to calculate payroll

- Learning is used when:

-Human expertise does not exist (like navigating on Mars),

- Humans are unable to explain their expertise (eg speech recognition)

- Solution changes in time (eg routing on a computer network)

- Solution needs to be adapted to particular cases (eg user biometrics)

2

Machine Learning – a few words

- Learning extracts general models from a data of particular examples

- Data is cheap and abundant (data warehouses, data marts etc) while knowledge is expensive and scarce.

- Example in retail: Customer transactions to consumer behavior:

  People who bought "Da Vinci Code" also bought "Angels and Demons" (www.amazon.com)

- Build a model that is a good and useful approximation to the data.

2 Machine Learning – a few words

- Study of algorithms that improve their performance
at some task with experience

- Optimize a performance criterion using example data or past experience.

- Role of Statistics: General conslucions from a sample

- Role of Computer science: Efficient algorithms to solve the optimization problem

- Representing and evaluating the model for useful conclusions

2 Machine Learning – a few words

Machine learning is preferred approach to

- Speech recognition, Natural language processing
- Computer vision
- Medical outcomes analysis
- Robot control
- Computational biology/Chemistry
- This trend is accelerating
- Improved machine learning algorithms
- Improved data capture, networking, faster computers

- Software too complex to write by hand
- New sensors / IO devices
- Demand for self-customization to user, environment
- When it turns out to be difficult to extract knowledge from human experts.

2 Machine Learning – a few words

Example applications include

- Association Analysis: Association analysis, also known as association rule mining or market basket analysis, is a Machine Learning technique used to discover interesting relationships, patterns, or associations among items in large datasets.

- Supervised Learning: Supervised learning involves training a model on a labeled dataset, where each data point is associated with a corresponding target label or outcome

- Unsupervised Learning: Unsupervised learning involves training a model on an unlabeled dataset, where the goal is to discover hidden patterns, structures, or relationships within the data.

- Reinforcement Learning: The agent learns through trial and error, receiving feedback in the form of rewards or penalties based on its actions.

2 Machine Learning –early example

- One common machine learning usage case in airport security is the development of systems to <u>track unattended luggage</u>.

- Unattended luggage poses a potential security risk in airports, as it could contain dangerous items or explosives.

- Machine learning algorithms can be utilized to help airport security personnel identify and monitor unattended luggage more effectively.

2 Machine Learning –early example

Key Features

- Video Surveillance: Many airports already have extensive video surveillance systems in place. Machine learning algorithms can be applied to analyze live video feeds from security cameras throughout the airport.

- Object Detection: Machine learning algorithms trained for object detection can be used to identify and track items such as luggage in the video feeds. These algorithms are trained on large datasets of images containing various objects, including luggage, and learn to recognize them based on their visual features.

2 Machine Learning –early example

Key Features

- Behavioral Analysis: Machine learning models can be trained to analyze the behavior of individuals in the airport. This includes factors such as how long someone has been in a certain area, how they move through the airport, and whether they appear to be acting suspiciously. Sudden abandonment of luggage or unusual behavior around luggage can trigger alerts for security personnel.

- Integration with Other Systems: Machine learning algorithms can be integrated with other security systems in the airport, such as access control systems and baggage handling systems. This allows for a more comprehensive approach to tracking unattended luggage and ensuring that any potential threats are addressed early on.

2 Machine Learning –early example

Key Features

- Alerting Security Personnel: When the machine learning system detects unattended luggage or suspicious behavior, it can automatically generate alerts for security personnel. These alerts can be sent to mobile devices carried by security personnel or displayed on monitoring screens in security control rooms.

- Continuous Learning: Machine learning systems can continuously learn and improve over time. By analyzing data on how security personnel respond to alerts and refining their algorithms based on this feedback, these systems can become more accurate and effective at identifying and tracking unattended luggage.

- Overall, by integrating machine learning algorithms with existing security systems, airports can improve their ability to detect and respond to security risks more effectively.

3 Big Data in Occupational Safety

- Big Data has numerous applications in occupational safety, helping organizations identify and mitigate risks, enhance worker well-being, and improve overall safety protocols.

- Implementing Big Data analytics in occupational safety not only enhances workplace safety but also contributes to a more proactive and data-driven approach to risk management.

3 Big Data in Occupational Safety

Examples include:

- Predictive Analytics for Accident Prevention: Analyzing historical data on workplace accidents can help identify patterns and trends, enabling predictive modeling to anticipate potential hazards and prevent accidents before they occur.

- Real-time Monitoring of Work Environments: Utilizing sensors and (Internet of Things) IoT devices to collect real-time data on various environmental factors such as temperature, air quality, noise levels, and radiation, which can help in identifying unsafe conditions and triggering immediate responses.

3 Big Data in Occupational Safety

- Wearable Technology for Health and Safety: Wearable devices can collect data on workers' vital signs, movements, and environmental exposure. Analyzing this data can help identify potential health risks, fatigue levels, and exposure to hazardous substances.

- Incident Investigation and Root Cause Analysis: Big Data analytics can assist in conducting thorough investigations into workplace incidents by analyzing data from various sources, including surveillance footage, witness statements, and equipment logs, to identify the root causes and prevent similar incidents in the future.

- Compliance Monitoring: Using Big Data to monitor and ensure compliance with safety regulations and standards by analyzing data from various sources, including inspections, audits, and incident reports.

3 Big Data in Occupational Safety

- Training and Simulation: Big Data can be used to analyze training data, identify areas of improvement, and customize safety training programs based on the specific needs and risks associated with different job roles and environments.

- Worker Behavior Analysis: Analyzing data on worker behavior can help identify unsafe practices and provide insights into the human factors contributing to accidents. This information can be used to design targeted interventions and training programs.

- Supply Chain Safety Monitoring: Tracking and analyzing data related to the safety performance of suppliers and contractors can help organizations assess and mitigate risks associated with their supply chain, ensuring that all parties involved adhere to safety standards.

3 Big Data in Occupational Safety

- Fatigue Management: Analyzing data on worker schedules, workload, and sleep patterns can help organizations identify and manage fatigue-related risks. This information can be crucial, especially in industries where alertness is paramount, such as transportation, healthcare, and manufacturing.

- Safety Culture Assessment: Big Data analytics can be employed to assess and measure the safety culture within an organization. By analyzing data related to safety reporting, employee surveys, and near-miss incidents, organizations can gain insights into the overall safety climate and make informed decisions to improve safety culture.

- Health Surveillance: Monitoring and analyzing health data, such as medical records, exposure to hazardous substances, and patterns of work-related illnesses, can assist in identifying emerging health risks and implementing preventive measures to protect workers.

3 Big Data in Occupational Safety

- Emergency Response Optimization: Big Data can improve emergency response plans by analyzing data on past incidents, simulating potential scenarios, and optimizing response strategies to minimize the impact of accidents and ensure the safety of workers.

- Continuous Improvement and Benchmarking: Using Big Data for ongoing performance monitoring and benchmarking against industry standards allows organizations to continuously improve their safety protocols and stay abreast of best practices.

Case Study 1 – Big Data- Occ.Safety

- Case Study: Predictive Analytics for Accident Prevention in Construction

- Background: A construction company, company A, implemented a Big Data analytics solution to enhance occupational safety on their construction sites. The company faced challenges with frequent accidents, and they aimed to proactively identify and address potential hazards to reduce the number of incidents.

- Objectives:

- Predict and prevent accidents by analyzing historical data.

- Improve overall safety performance and compliance with regulations.

- Enhance decision-making for resource allocation and safety protocols.

Implementation:

- Data Collection:

Company A collected a vast amount of data, including:

- Historical accident records.

- Weather conditions during construction activities.

- Employee work hours and schedules.

- Equipment usage and maintenance logs.

- Site layouts and blueprints.

3 Case Study 1 – Big Data- Occ.Safety

Integration of IoT Sensors:

- The company integrated IoT sensors on construction equipment, personal protective equipment (PPE), and throughout the construction site to collect real-time data.

- Equipment sensors monitored usage patterns, identifying potential malfunctions.

- PPE sensors tracked workers' movements, vital signs, and exposure to environmental conditions.

3 Case Study 1 – Big Data- Occ.Safety

Predictive Analytics Model:

- Data scientists at Company A developed a predictive analytics model using machine learning algorithms. The model analyzed historical data to identify patterns and correlations associated with past accidents.

- Factors such as weather conditions, time of day, equipment usage, and worker fatigue were considered in predicting potential risks.

Case Study 1 – Big Data- Occ.Safety

Real-time Monitoring Dashboard:

- A centralized dashboard was created to provide real-time insights to safety officers and project managers.

- The dashboard displayed:

- Predictive risk scores for different areas of the construction site.

- Alerts for conditions that indicated a higher risk of accidents.

- Recommendations for preventive actions based on the predictive model's outputs.

3

Case Study 1 – Big Data- Occ.Safety

Intervention and Preventive Measures:

- When the predictive model identified elevated risk levels, safety officers received immediate alerts. They could then take proactive measures such as:

- Temporarily halting specific operations.

- Adjusting work schedules or reallocating resources.

- Conducting targeted safety training for workers in high-risk areas.

Case Study 1 – Big Data- Occ.Safety

Results:

■ Accident Reduction: Company A observed a significant reduction in the number of accidents on construction sites after implementing the predictive analytics solution.

■ Resource Optimization: The company could allocate resources more efficiently by focusing on high-risk areas, leading to improved productivity and cost savings.

■ Proactive Safety Measures: The ability to predict potential hazards allowed for proactive interventions, fostering a culture of safety and reducing the likelihood of severe incidents.

3

Case Study 1 – Big Data- Occ.Safety

Lessons Learned:

- The integration of real-time data and predictive analytics provided actionable insights for immediate decision-making.

- Continuous monitoring and refinement of the predictive model were essential for adapting to changing site conditions and improving accuracy over time.

- Employee engagement and training were crucial for the success of the initiative, as workers needed to understand and embrace the new safety protocols.

- This case study demonstrates how Big Data analytics, particularly predictive modeling, can be a powerful tool in enhancing occupational safety by identifying and mitigating risks before accidents occur.

3 Case Study 2 – Big Data- Occ.Safety

Case Study: Wearable Technology for Worker Safety in Manufacturing

- Background: Company B, a large-scale production facility, aimed to improve worker safety and reduce the occurrence of workplace injuries. The company decided to leverage wearable technology and Big Data analytics to monitor and enhance the well-being of its workforce.

- Objectives:

- Monitor workers' health and safety in real-time.

- Identify patterns related to fatigue, ergonomics and environmental factors.

- Implement preventive measures to reduce the risk of injuries.

Implementation:

- Wearable Devices Deployment: Company B provided workers with wearable devices equipped with sensors to monitor various parameters.

- Accelerometers tracked body movements and postures.

- Heart rate monitors and temperature sensors tracked physiological data.

- GPS modules recorded the location and movement of workers within the facility.

- Data Integration and Storage:

- The data collected from wearable devices were integrated into a centralized database.

- Additional data sources, such as production schedules, environmental conditions, and historical safety records, were also incorporated for a comprehensive analysis.

Case Study 2 – Big Data- Occ.Safety

- Machine Learning Algorithms: Machine learning algorithms were applied to analyze the integrated data and identify patterns associated with potential safety risks.

Algorithms were trained to recognize correlations between certain work conditions, physical exertion, and the likelihood of fatigue or ergonomic issues.

- Real-time Monitoring and Alerts: A real-time monitoring system was implemented to provide instant feedback to workers and supervisors.

If the algorithms detected signs of fatigue, unsafe postures, or exposure to hazardous conditions, immediate alerts were sent to both the worker and their supervisor.

- Customized Training Programs: The insights derived from the Big Data analytics were used to develop personalized safety training programs.

Workers received targeted training on ergonomics, proper lifting techniques, and the importance of taking breaks to prevent fatigue.

3 Case Study 2 – Big Data- Occ.Safety

Results:

- Injury Reduction: Company B experienced a notable reduction in workplace injuries as a result of early detection and intervention.

- Improved Ergonomics: The data-driven insights led to the redesign of workstations and processes, improving ergonomic conditions and reducing the risk of musculoskeletal injuries.

- Enhanced Productivity: By addressing fatigue and ergonomic issues, workers reported increased comfort and focus, contributing to improved overall productivity.

3 Case Study 2 – Big Data– Occ.Safety

Lessons Learned:

- Wearable technology provided real-time, detailed data, allowing for more accurate risk assessments.

- The integration of multiple data sources provided a holistic view of the work environment and its impact on worker safety.

- Continuous monitoring and feedback were essential for the success of the program, allowing for quick adjustments to safety protocols based on evolving conditions.

- This case study highlights how the combination of wearable technology and Big Data analytics can play a crucial role in promoting worker safety by providing real-time insights and enabling proactive measures to mitigate risks.

4a

A few words about Computational Chemistry

- Computational chemistry is a multidisciplinary field that utilizes computational methods and theoretical models to understand and predict the behavior of molecules and chemical reactions. It combines principles from chemistry, physics, mathematics, and computer science to simulate and analyze molecular systems at the atomic level.

- By employing computational techniques such as molecular dynamics simulations, quantum mechanical calculations, and molecular modeling, computational chemistry provides insights into a wide range of phenomena, including chemical reactions, molecular properties, protein-ligand interactions, and material properties.

- This field plays a crucial role in various scientific disciplines and industries, including drug discovery, materials science, catalysis, environmental chemistry, and biochemistry. Computational chemistry enables researchers to explore complex molecular systems, design novel materials and drugs, optimize chemical processes, and contribute to the advancement of scientific knowledge and technological innovation.

4a A few words about Computational Chemistry - Methods

- Molecular Mechanics (MM): Molecular mechanics methods simplify molecular systems by representing atoms as classical particles interacting through empirical force fields. These force fields describe the energy associated with bond stretching, angle bending, torsional rotations, and non-bonded interactions (such as van der Waals and electrostatic interactions). Molecular mechanics calculations are often used for energy minimization, molecular dynamics simulations, and structure optimization of large biomolecules and chemical systems.

- Quantum Mechanics (QM): Quantum mechanics methods solve the Schrödinger equation to describe the electronic structure and behavior of atoms and molecules at the quantum level. These methods range from simple approaches like Hartree-Fock theory to more advanced techniques such as density functional theory (DFT), Møller-Plesset perturbation theory (MP2), and coupled cluster theory (CC). Quantum mechanics calculations provide accurate descriptions of molecular properties, electronic energies, and reaction mechanisms, making them essential for studying chemical bonding, spectroscopy, and reaction kinetics.

4a

A few words about Computational Chemistry

- Density Functional Theory (DFT): DFT is a widely used quantum mechanical method that approximates the electronic density of a system rather than solving the full many-body Schrödinger equation. DFT calculations are computationally efficient and provide accurate predictions of molecular structures, energies, and properties. They are commonly employed for studying molecular energetics, reaction mechanisms, and material properties.

- Molecular Dynamics (MD): Molecular dynamics simulations simulate the time evolution of molecular systems by numerically integrating the equations of motion for all atoms in the system. MD methods use classical or quantum mechanical force fields to describe the interactions between atoms and molecules. Molecular dynamics simulations provide insights into the dynamic behavior, conformational changes, and thermodynamic properties of biomolecules, materials, and chemical reactions.

4a A few words about Computational Chemistry

- Monte Carlo (MC) Methods: Monte Carlo methods use random sampling techniques to explore the configuration space of molecular systems and calculate thermodynamic properties such as free energy, entropy, and partition functions. MC simulations can be employed to study phase transitions, molecular assemblies, and equilibrium properties of complex systems.

- Hybrid Methods: Hybrid methods combine different computational approaches, such as QM/MM (quantum mechanics/molecular mechanics) and QM/QM' (quantum mechanics/quantum mechanics'), to model molecular systems with varying levels of accuracy and computational cost. These methods enable the study of large biomolecular complexes, enzyme reactions, and catalytic mechanisms by treating different regions of the system at different levels of theory.

4 Big Data in Chemistry

- Scientists have long benefitted from and contributed to the development of quantitative methods to reveal patterns in structure-property relationships across all branches of chemistry ranging from materials to synthetic organic to biological.

- Recent advances in computing power, software and algorithms, as well as increases in data availability from experiment or computation, have led to dramatic progress in the complexity of statistical techniques applied to chemistry.

4   Big Data in Chemistry

Big Data applications in chemistry have the potential to revolutionize research, drug discovery, and various processes within the field. A few applications of Big Data in chemistry are:

- Drug Discovery and Development: Analyzing vast datasets related to chemical structures, biological interactions, and pharmacological profiles can expedite the drug discovery process. Big Data helps identify potential drug candidates, predict their efficiency, and optimize molecular structures for improved bioavailability.

- Cheminformatics and Virtual Screening: Big Data tools enable the analysis of chemical databases to identify patterns, similarities, and relationships between different compounds. Virtual screening techniques use these insights to predict the bioactivity of molecules, aiding in the selection of promising candidates for further experimental validation.

4 Big Data in Chemistry

- Materials Science and Design: Big Data analytics can be employed to explore the vast space of materials properties, enabling the design of novel materials with specific characteristics. This is crucial for applications in electronics, energy storage, catalysis, and other areas of Material Science.

- Quantum Chemistry Simulations: Processing large datasets generated by quantum chemistry simulations allows researchers to model and understand complex molecular interactions at a quantum level. This aids in the exploration of reaction mechanisms, electronic structures, and properties of materials.

- Chemical Informatics for Toxicology and Safety Assessment: Big Data analytics help assess the toxicity and safety of chemical compounds by integrating information from diverse sources, including chemical databases, biological assays, and toxicology studies. This contributes to a more comprehensive understanding of the potential risks associated with specific substances.

4 Big Data in Chemistry

- Metabolomics and Systems Biology: Analyzing large-scale metabolomic data helps researchers understand the metabolic pathways within organisms. Big Data applications in metabolomics contribute to identifying biomarkers, understanding disease mechanisms, and developing personalized medicine approaches.

- Analytical Chemistry and Spectroscopy: Big Data tools assist in the analysis of complex spectroscopic data, such as mass spectrometry and nuclear magnetic resonance (NMR). These techniques generate vast datasets that can be processed to identify chemical compounds, assess purity, and characterize molecular structures.

- Pharmacogenomics and Personalized Medicine: Integrating genetic, clinical, and chemical data enables the development of personalized medicine approaches. Big Data applications in pharmacogenomics (how a person's genetic profile affects how their body processes medicine) help identify genetic variations that influence individual responses to drugs, leading to more tailored and effective treatments.

4 Big Data in Chemistry

- Chemical Reaction Prediction: Predictive modeling using Big Data techniques can assist in predicting chemical reactions and their outcomes. This is valuable for optimizing reaction conditions, exploring synthesis pathways, and designing efficient chemical processes.

- Environmental Monitoring and Chemistry: Big Data is applied in environmental chemistry to analyze large datasets related to air and water quality, pollution levels, and the impact of chemical substances on ecosystems. This information aids in environmental monitoring, risk assessment, and regulatory compliance.

- These applications highlight how Big Data in chemistry can transform the way researchers explore, analyze, and apply chemical knowledge, ultimately leading to advancements in drug development, materials science, and our understanding of molecular processes.

4 Big Data in Computational Chemistry

Big Data applications in computational chemistry leverage large datasets and advanced analytics to enhance the understanding of molecular structures, predict chemical properties, and accelerate drug discovery. Here are a few applications in computational chemistry:

- Molecular Dynamics Simulations: Big Data techniques are employed to analyze and interpret the vast amount of data generated from molecular dynamics simulations. This allows researchers to study the behavior of molecules over time, explore conformational changes, and gain insights into the dynamics of biological macromolecules.

- Quantum Chemistry Calculations: Big Data analytics can process and analyze the results of quantum chemistry calculations. This includes predicting electronic structures, molecular properties, and reaction mechanisms at a quantum level. Handling the extensive datasets generated by these simulations is crucial for understanding complex molecular interactions.

- Chemical Database Mining: Computational chemistry relies on the analysis of chemical databases for structure-activity relationships, compound screening, and virtual screening. Big Data tools help in efficient mining of large chemical databases to identify patterns, similarities, and trends among molecular structures.

4 Case Study: Big Data in Computational Chemistry

- Case Study: IBM's World Community Grid for Drug Discovery

- Background: IBM's World Community Grid is a distributed computing platform that utilizes the power of volunteers' idle computing resources worldwide to contribute to scientific research. One notable project on this platform is "Help Cure Muscular Dystrophy" (HCMDB@Home), which involves extensive computational chemistry simulations for drug discovery.

- Objectives:

- Leverage the computational power of volunteers' devices to perform large-scale molecular dynamics simulations.

- Analyze vast datasets to identify potential drug candidates for muscular dystrophy.

4 Case Study: Big Data in Computational Chemistry

- Implementation:

- Molecular Dynamics Simulations: The project focused on simulating the behavior of proteins associated with muscular dystrophy using molecular dynamics simulations.

- These simulations aimed to understand the dynamic interactions between proteins, identify potential binding sites, and explore the stability of protein structures.

- Distributed Computing:

- Utilizing IBM's World Community Grid, volunteers from around the world installed a screensaver application that ran molecular dynamics simulations during their devices' idle time.

- The collective computational power of thousands of devices allowed for the generation of massive amounts of simulation data.

4 Case Study: Big Data in Computational Chemistry

- Big Data Analytics:

- The project generated large datasets containing information about protein conformations, interactions, and dynamics.

- Big Data analytics tools were employed to analyze this wealth of data, identifying patterns and potential drug binding sites.

- Virtual Screening and Drug Discovery:

- The analyzed data were used for virtual screening, a process that involves computationally testing a vast number of small molecules to identify potential drug candidates.

- Big Data tools facilitated the screening process, helping prioritize compounds based on their predicted interactions with the target proteins.

- Iterative Optimization:

- The drug discovery process involved iterative optimization, where the computational simulations were continually refined based on the insights gained from the analysis of previous simulation data.

- This iterative approach allowed researchers to improve the accuracy of predictions and enhance the understanding of the molecular mechanisms involved in muscular dystrophy.

4 Case Study: Big Data in Computational Chemistry

Results:

- The project on IBM's World Community Grid significantly accelerated the drug discovery process for muscular dystrophy.

- Thousands of volunteers collectively contributed vast computational resources, enabling extensive simulations that would have been impractical with traditional computing resources.

- The project identified potential drug candidates and provided valuable insights into the molecular mechanisms underlying muscular dystrophy.

4 Case Study: Big Data in Computational Chemistry

Impact and Lessons Learned:

- The case demonstrates the power of distributed computing and Big Data analytics in advancing drug discovery for rare diseases.

- By engaging a global community of volunteers, the project showcased the potential of collaborative, crowd-sourced computing for scientific research.

- The success of this project encouraged the exploration of similar approaches for other diseases and research areas, highlighting the scalability and impact of Big Data applications in computational chemistry.

- This case study illustrates how the combination of distributed computing and Big Data analytics on a global scale can significantly contribute to accelerating drug discovery efforts in the field of computational chemistry.

5 Machine Learning in Occ. Safety

Machine Learning (ML) plays a significant role in improving occupational safety by analyzing large datasets, identifying patterns, and predicting potential risks. Here are several applications of Machine Learning in occupational safety:

- Risk Prediction and Prevention: ML algorithms can analyze historical safety data, incident reports, and environmental conditions to predict and prevent workplace accidents. By identifying patterns and correlations, the system can provide early warnings about potential safety risks.

- Behavioral Safety Analysis: ML models can analyze workers' behavior data, including their movements and actions captured by wearable devices or surveillance cameras. This helps in identifying unsafe behaviors and providing targeted interventions or training to improve safety practices.

5 Machine Learning in Occ. Safety

- Predictive Maintenance: ML can predict equipment failures by analyzing data from sensors and maintenance records. This proactive approach allows organizations to address potential safety hazards related to malfunctioning machinery before they cause accidents.

- Safety Compliance Monitoring: ML algorithms can assess compliance with safety regulations by analyzing data from inspections, audits, and incident reports. This helps organizations ensure that safety protocols are followed and identify areas that require corrective actions.

- Fatigue Detection: ML models can analyze data from wearable devices or monitoring systems to detect signs of worker fatigue. By identifying patterns in movement, heart rate, and other physiological indicators, the system can issue alerts when fatigue levels increase, helping prevent accidents related to impaired alertness.

5 Machine Learning in Occ. Safety

- Emergency Response Optimization: ML algorithms can optimize emergency response plans by analyzing historical data on incidents. This includes predicting the likely impact of emergencies, identifying evacuation routes, and optimizing resource allocation for a more effective response.

- Personalized Safety Training: ML can tailor safety training programs based on individual worker profiles, job roles, and historical safety data. Personalized training helps address specific safety concerns and improves the overall safety knowledge of employees.

- Near-Miss Analysis: ML can analyze data related to near-miss incidents to identify patterns and root causes. Understanding near misses helps organizations proactively address potential hazards and prevent future accidents.

5 Machine Learning in Occ. Safety

- Occupational Health Monitoring: ML models can monitor and analyze health data from workers, including physiological parameters and exposure to occupational hazards. This helps in identifying health risks and implementing preventive measures to protect the well-being of employees.

- Smart Personal Protective Equipment (PPE): ML can be used to enhance the effectiveness of PPE by integrating sensors that monitor environmental conditions and alert workers to potential dangers. This ensures that PPE is used appropriately and provides an additional layer of safety.

- Workforce Safety Analytics: ML-driven analytics platforms can aggregate and analyze various safety-related data sources, providing organizations with insights into overall safety performance, trends, and areas that need improvement.

- Collaborative Robotics Safety: ML algorithms contribute to the safe collaboration between humans and robots in the workplace. These algorithms can continuously analyze sensor data to ensure that robots operate safely around human workers.

5

Case Study: Machine Learning in Occupational Safety

Case Study: Predictive Maintenance for Workplace Equipment Safety

- Background: A manufacturing company, Company C, implemented a predictive maintenance system powered by Machine Learning to enhance workplace equipment safety. The company aimed to proactively identify potential issues with machinery and equipment that could pose safety risks to workers.

- Objectives:

- Predict equipment failures before they occur to prevent accidents and downtime.

- Improve overall workplace safety by addressing maintenance needs promptly.

- Optimize resource allocation and extend the lifespan of machinery.

5

Case Study: Machine Learning in Occupational Safety

Implementation:

- Data Collection:

- Company C gathered historical data on equipment performance, maintenance records, and incidents related to equipment failures.

- Sensor data from the machines, including temperature, vibration, and other relevant parameters, was collected to monitor real-time conditions.

- Feature Engineering:

- Relevant features were extracted from the collected data to create a comprehensive dataset for model training.

- Features included equipment runtime, historical maintenance patterns, sensor readings, and environmental conditions.

5 Case Study: Machine Learning in Occupational Safety

- Machine Learning Model Development:

- A predictive maintenance model was developed using Machine Learning algorithms, such as regression or classification models, depending on the nature of the prediction task.

- The model was trained on historical data to learn patterns associated with equipment failures and maintenance needs.

- Real-time Monitoring and Alerts:

- The Machine Learning model was integrated into a real-time monitoring system that continuously analyzed sensor data from the equipment.

- When the model detected patterns indicative of potential equipment failure or the need for maintenance, it triggered immediate alerts for the maintenance team.

- Dynamic Maintenance Scheduling:

- The system dynamically adjusted maintenance schedules based on the predictions from the Machine Learning model.

- High-priority maintenance tasks were scheduled promptly, reducing the likelihood of unexpected equipment failures and associated safety risks.

5 Case Study: Machine Learning in Occupational Safety

Results:

- Accident Prevention: By predicting equipment failures in advance, Company C significantly reduced the risk of accidents caused by malfunctioning machinery.

- Downtime Reduction: Proactive maintenance based on Machine Learning predictions led to a decrease in unplanned downtime, enhancing overall productivity.

- Resource Optimization: The company optimized the allocation of maintenance resources by focusing on equipment with the highest likelihood of failure, reducing unnecessary maintenance on well-functioning machinery.

5 Case Study: Machine Learning in Occupational Safety

Lessons Learned:

- Data Quality is Crucial: High-quality data is essential for training accurate Machine Learning models. Ensuring accurate and timely data collection from sensors and maintenance records is crucial for the success of the predictive maintenance system.

- Continuous Monitoring and Iterative Improvement: The predictive maintenance system requires continuous monitoring and iterative improvement. Regular updates to the Machine Learning model based on new data and evolving conditions contribute to its effectiveness over time.

- Integration with Existing Systems: Integrating the predictive maintenance system with existing workplace safety and maintenance systems allows for a seamless workflow and ensures that safety protocols are aligned with maintenance activities.

- This case study illustrates how Machine Learning, in the context of predictive maintenance, contributes to occupational safety by preventing equipment failures and minimizing risks associated with machinery malfunctions in the workplace.

6 Machine Learning in Computational Chemistry

Machine Learning (ML) has found various applications in the field of computational chemistry, aiding researchers in predicting molecular properties, optimizing chemical processes, and accelerating drug discovery. Here are a few applications of Machine Learning in computational chemistry:

- Quantum Chemistry Calculations: ML models can be trained to predict quantum mechanical properties of molecules. This includes predicting molecular energy levels, electronic structures, and other quantum descriptors, allowing for faster and more efficient quantum chemistry simulations.

- Molecular Property Prediction: ML algorithms can predict various molecular properties, such as solubility, toxicity, and bioavailability. This information is valuable in drug discovery and materials science for selecting compounds with desired properties.

6 Machine Learning in Computational Chemistry

- Drug Discovery and Virtual Screening: ML is widely used in virtual screening to predict the binding affinity of small molecules to biological targets. These models help identify potential drug candidates, prioritize compounds for experimental testing, and optimize lead compounds for better efficacy.

- Chemical Reaction Prediction: ML models can predict the outcomes of chemical reactions, including the products and reaction conditions. This assists in designing synthetic routes, optimizing reaction conditions, and exploring new chemical transformations.

- QSPR (Quantitative Structure-Property Relationship) Modeling: ML is applied in building QSPR models to correlate the molecular structure of compounds with their physical and chemical properties. This helps in predicting properties such as boiling points, melting points, density etc.

6

Machine Learning in Computational Chemistry

- Materials Design and Discovery: ML facilitates the prediction of material properties by analyzing data on the structure and composition of materials. This is valuable in designing new materials with specific characteristics for applications in electronics, energy storage, and catalysis.

- Protein-Ligand Binding Prediction: ML models can predict the binding affinity and interactions between proteins and ligands. This is crucial in understanding the mechanisms of drug-target interactions and aiding in the design of more effective therapeutic compounds.

- Conformational Analysis: ML techniques are applied to analyze and predict the conformational flexibility of molecules. Understanding molecular conformations is essential for studying the behavior of biomolecules and optimizing drug candidates.

6 Machine Learning in Computational Chemistry

- Chemical Informatics for High-Throughput Screening: ML is employed in cheminformatics to analyze high-throughput screening data. This includes the identification of active compounds, understanding structure-activity relationships, and prioritizing compounds for further experimental validation.

- Spectral Analysis and NMR Prediction: ML models can predict nuclear magnetic resonance (NMR) spectra and other spectral data based on molecular structures. This aids in the interpretation of experimental data and accelerates the process of structure elucidation.

- Reaction Pathway Prediction: ML algorithms can predict reaction pathways by analyzing data on known reactions. This helps in proposing plausible reaction mechanisms and designing synthetic routes for chemical synthesis.

- Scaffold Hopping in Drug Design: ML assists in scaffold hopping, a process that explores diverse molecular frameworks to design novel drug-like compounds. This allows researchers to identify alternative scaffolds with similar biological activity.

6 Case Study 1: Machine Learning in Computational Chemistry

Case Study: Predictive Modeling for Molecular Property Prediction

- Background: A pharmaceutical research company, Company D, implemented a Machine Learning-based predictive modeling approach to accelerate the drug discovery process. The focus was on predicting key molecular properties that impact drug efficacy and safety.

- Objectives:

- Accelerate the identification of potential drug candidates by predicting relevant molecular properties.

- Reduce the reliance on time-consuming and resource-intensive laboratory experiments.

- Enhance the efficiency of the drug discovery pipeline through data-driven decision-making.

6 Case Study 1: Machine Learning in Computational Chemistry

Implementation:

- Data Collection and Preparation:

- Company D compiled a comprehensive dataset comprising chemical structures of known compounds, along with corresponding experimental data on molecular properties.

- Data preprocessing involved cleaning, standardization, and feature extraction to prepare the dataset for training Machine Learning models.

- Feature Engineering:

- Relevant molecular descriptors and features were extracted from the chemical structures, including 2D and 3D descriptors, physicochemical properties, and structural fingerprints.

- Additional features related to the biological activity and toxicity of compounds were included to capture a broader range of information.

6 Case Study 1: Machine Learning in Computational Chemistry

- Machine Learning Model Selection:

- Various Machine Learning algorithms, including regression models, random forests, and neural networks, were evaluated for their ability to predict specific molecular properties.

- Cross-validation techniques were employed to assess model performance and select the most suitable algorithm for each property.

- Model Training and Validation:

- The selected Machine Learning models were trained on the prepared dataset, learning the relationships between chemical structures and molecular properties.

- Validation was performed using separate datasets to ensure that the models could generalize well to new, unseen compounds.

- Integration into Drug Discovery Pipeline:

- The trained Machine Learning models were integrated into Company D drug discovery pipeline.

- Computational chemists and medicinal chemists used the predictive models to prioritize compounds for synthesis and experimental testing based on their predicted molecular properties.

6 Case Study 1: Machine Learning in Computational Chemistry

Results:

- Accelerated Drug Discovery: The Machine Learning models significantly accelerated the identification of potential drug candidates by predicting key molecular properties.

- Resource Optimization: Company D reduced the need for extensive laboratory experiments by strategically selecting compounds with favorable predicted properties for further testing.

- Improved Hit-to-Lead Process *Hit-to-Lead is the drug discovery process of synthetizing and testing many compounds to determine the most promising): The predictive models enhanced the hit-to-lead optimization process by guiding researchers towards chemical modifications that were more likely to improve desired molecular properties.

6 Case Study 1: Machine Learning in Computational Chemistry

Lessons Learned:

- Quality of Training Data: The success of the predictive models relied on the quality and representativeness of the training data. Ensuring diverse and comprehensive data improves the model's ability to generalize to new compounds.

- Iterative Model Refinement: Continuous feedback loops were established to iteratively refine the Machine Learning models based on new experimental data and insights from the drug discovery process.

- Interdisciplinary (many different areas) Collaboration: Effective collaboration between computational chemists, data scientists, and experimental researchers was crucial for the successful integration of Machine Learning models into the drug discovery pipeline.

6 Case Study 2: Machine Learning in Computational Chemistry

Problem: Predicting the Mechanical Properties of Metal Alloys

- Background: Metal alloys are widely used in various industries for their desirable mechanical properties, such as strength, ductility, and toughness. However, designing new metal alloys with optimized mechanical properties often requires extensive experimental testing, which can be time-consuming and costly.

- Objective: Develop a predictive model using Machine Learning and Molecular Dynamics simulations to estimate the mechanical properties of metal alloys based on their atomic structure

6 Case Study 2: Machine Learning in Computational Chemistry

Approach:

- Data Collection: Gather a dataset comprising atomic structures of different metal alloys along with their corresponding mechanical properties, such as Young's modulus, yield strength, and fracture toughness. These atomic structures can be obtained from experimental observations, literature databases, or computational simulations.

- Feature Engineering: Extract relevant features from the atomic structures of metal alloys that may influence their mechanical properties. These features may include atomic composition, crystal structure, grain boundaries, defects, and interatomic distances.

6 Case Study 2: Machine Learning in Computational Chemistry

- Machine Learning Model Development:

- Train a Machine Learning model, such as a <u>regression model or neural network</u>, using the dataset of atomic structures and mechanical properties.

- Utilize Molecular Dynamics simulations to generate additional training data by simulating the mechanical behavior of metal alloys under various loading conditions.

- Incorporate physical principles and domain knowledge into the feature selection and model development process to improve the accuracy and interpretability of the model.

6 Case Study 2: Machine Learning in Computational Chemistry

- Model Validation and Evaluation: Validate the Machine Learning model using cross-validation techniques or by splitting the dataset into training and testing subsets. Evaluate the model's performance in predicting mechanical properties on unseen data and compare it with experimental results or high-fidelity simulations.

- Model Deployment and Prediction: Once validated, deploy the trained Machine Learning model to predict the mechanical properties of new metal alloys based on their atomic structure. This predictive model can be used to screen and prioritize candidate alloys for further experimental testing or optimization.

6

Case Study 2: Machine Learning in Computational Chemistry

Benefits:

- Accelerated Materials Design: The predictive model enables rapid screening of a large number of metal alloys, accelerating the materials design process and reducing the need for extensive experimental testing.

- Insight into Structure-Property Relationships: Machine Learning combined with Molecular Dynamics simulations provides insights into the underlying atomic-scale mechanisms governing the mechanical behavior of metal alloys, aiding in the design of new materials with tailored properties.

- Cost and Resource Savings: By reducing the reliance on experimental testing, the predictive model saves time, resources, and costs associated with materials development, leading to more efficient and cost-effective materials design strategies.

Thanks for your attention

...and patience.

Questionnaire will be available onwards ………………………………………………..