CSCE 590-1: From Data to Decisions with Open Data: A Practical Introduction to AI

Lecture 1: Introductions: AI, Data, Decisions

PROF. BIPLAV SRIVASTAVA, AI INSTITUTE

12TH JAN 2021

Carolinian Creed: "I will practice personal and academic integrity."

Organization of Lecture 1

- Introduction Segment
  - Course Logistics
- Main Segment
  - AI: A Quick Introduction
  - Data
  - Decisions
- Concluding Segment
  - About Next Lecture – Lecture 2
  - Ask me anything

Course Logistics

Administrative Information

- Office Hours: W,F 11:30 am – 12:30 pm or by appointment
- Websites
  - Course: https://blackboard.sc.edu
  - Supplementary:
  - https://sites.google.com/site/biplavsrivastava/teaching/csce-590-from-data-to-decisions-with-open-data-a-practical-introduction-t?authuser=0
- Class methods
  - Asynchronous Online: Blackboard
  - In-class: TBD, only when feasible

Learning Objectives

Undergraduate students will be able to:

L1: **Identify patterns** in problems around us that can be solved with better information / insights derived from data. Example: gap in information about demand and supply.

L2: Explain opportunities, **issues related to data and tools**: (a) data meant for reuse, i.e., open data (b) data quality, (c) data integration, (d) privacy concerns and bias with data, (e) experiment design,

L3: Explain, execute and **create analytical methods to process data**: (a) unstructured data, (b) semi-structured data, (c) structured data

L4: Explain **AI methods in data analysis**: (a) Learning methods, (b) Reasoning, (c) Representation and standardization – knowledge graphs/ ontology, (d) Preferences, (e) Handling Uncertainty

Graduate students will be able to do all of the above, and:

L5: Evaluate **gaps in analytical methods** and create new ones to process data

L6: **Explain data-driven insights to end-users** with user-oriented interfaces, provide explanations for produced output to build trust. Using interactive interfaces, like visualizations and chatbots, explain how users will be able to interact with insights and build trust in AI.

L7: Explain **research findings in open areas** and critique their contributions

Course Material

- Python for Data Analysis
  - Latest: Python for Data Analysis Book, by Wes McKinney, 2nd Edition. On Amazon at: https://www.amazon.com/gp/product/1491957662/, ISBN-13: 978-1491957660, ISBN-10: 1491957662
  - Book Data and Code Notebooks: https://github.com/wesm/pydata-book
  - 1st edition (free download): https://bedford-computing.co.uk/learning/wp-content/uploads/2015/10/Python-for-Data-Analysis.pdf
- Artificial Intelligence: A Modern Approach (Fourth edition, 2020), Stuart Russell and Peter Norvig,
  - http://aima.cs.berkeley.edu/, ISBN-13: 978-0134610993
- Open Datasets
  - US: https://www.data.gov/ or any US state
  - Text of legislations - LegiScan, https://legiscan.com/
  - Kaggle datasets: https://www.kaggle.com/datasets
  - Google datasets search: https://datasetsearch.research.google.com/

Undergraduate Student Assessment

Tests	1000 points
• Course Project – report, in-class presentation	600 points
• Quiz – best of 3 from 4	200 points
• Final Exam	200 points
Total	1000 points

- Project: 50% + 10%: project report (50%) and code, for elevator presentation to class (10%)
  - Data analysis project
    - Dataset must be from given catalog
    - Use analytical methods to present new insights
- Quiz: 20%
  - 4 based on preceding lectures
- Exam: 20%
  - For undergraduate, final examination. Total 20%

Tests	1000 points
• Course Project – report, in-class presentation	600 points
• Quiz – best of 3 from 4	200 points
• Final Exam – Paper summary, in-class presentation	200 points
Total	1000 points

Graduate Student Assessment

- Project: 50% + 10%: project report (50%) and code, for elevator presentation to class (10%)
  - Data analysis project OR
    - Dataset must be from given catalog
    - Use analytical methods to present new insights
  - Create or explore new methods (preferred for graduate students) project
    - Problem to be discussed with instructor
    - Example: Analyze sound signals to estimate crowd
- Quiz: 20%
  - 4 based on preceding lectures
- Exam:
  - Research paper reading (10%) and presentation to class (10)% - Total 20%
    - Read a paper accepted at a top Data / AI conference: AAAI 2019-2021, IJCAI 2019-2021, NeurIPS 2019-2021, KDD 2019-2021, SIGMOD 2019-2021. Make a 1-page summary highlighting the key points, what you liked and what you did not. Try any code given in the paper
    - Present a 1-slide summary to class (10%)

Student Assessment

A = [900-1000]
B+ = [870-899]
B = [800-869]
C+ = [770-799]
C = [700-769]
D+ = [670-699]
D = [600-669]
F = [0-599]

Tests	1000 points
• Course Project	600 points
• Quiz	200 points
• Final Exam	200 points
Total	1000 points

Motivating Scenarios

Example: Courses for a Student

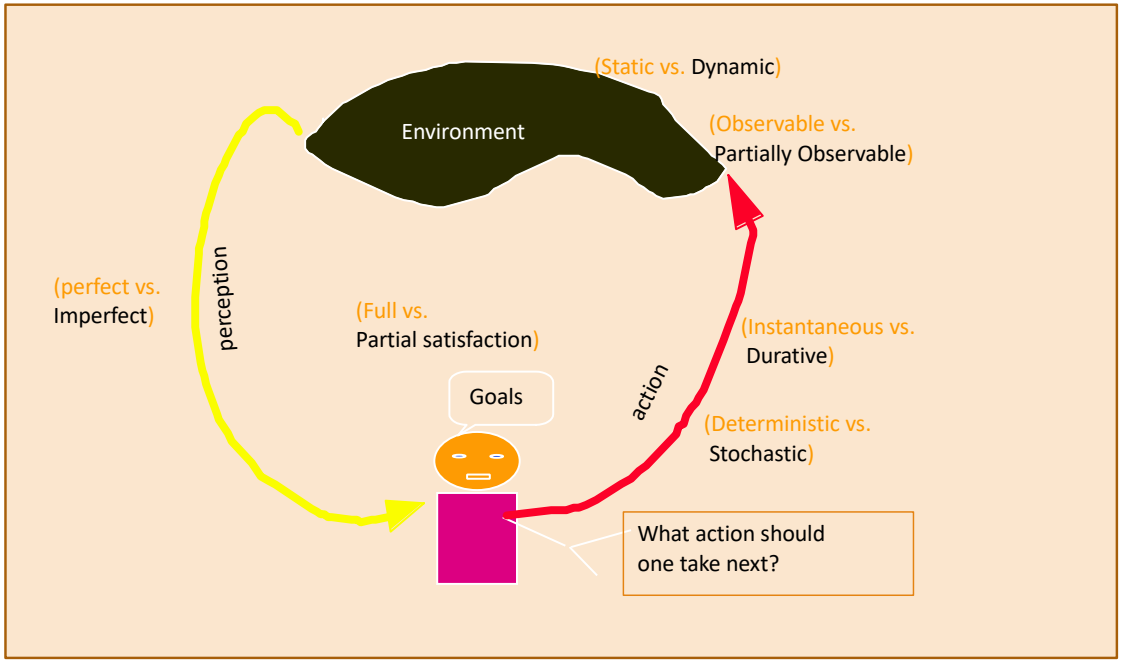
- Decision: Student deciding which courses to take for their program
- Data
  - About courses
  - About faculties
  - About job opportunities
  - About research opportunities and industry trends
  - Private: what the student wants to do
- Analysis
  - Courses offered in different semesters
  - Teachers offering courses – background, hardness of classes, …

Example: Health During a Pandemic

- Decision: Individual staying healthy during a pandemic like COVID19
- Data
  - About cases
  - About mitigation steps: e.g., mask wearing restrictions and practices, lockdowns, hospital conditions
  - Private: pre-existing health conditions
- Analysis
  - Regions with high and low cases
  - Whether to eat inside a restaurant?
  - How to make an urgent road trip ?
  - How to hold classes at a University?

AI: A Quick Introduction

Artificial Intelligence (AI) as an Agent

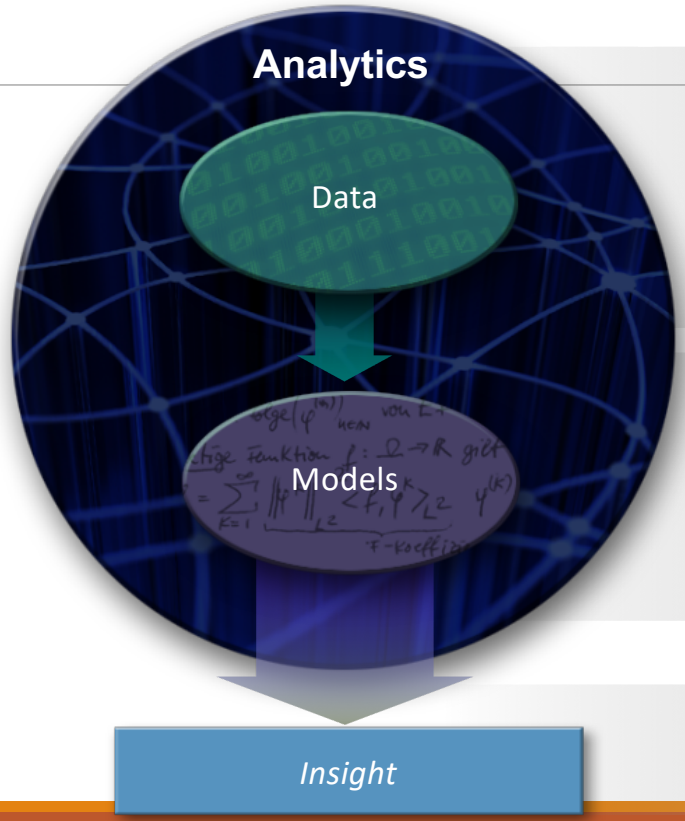


AI deals with perceiving the environment and taking actions towards short- and long term goals as the world changes over time.

From Subbarao Kambhampati's AI Planning Course

Advanced AI Techniques (Analytics) like Reasoning & Machine Learning

make use of data and models to provide insight to guide decisions

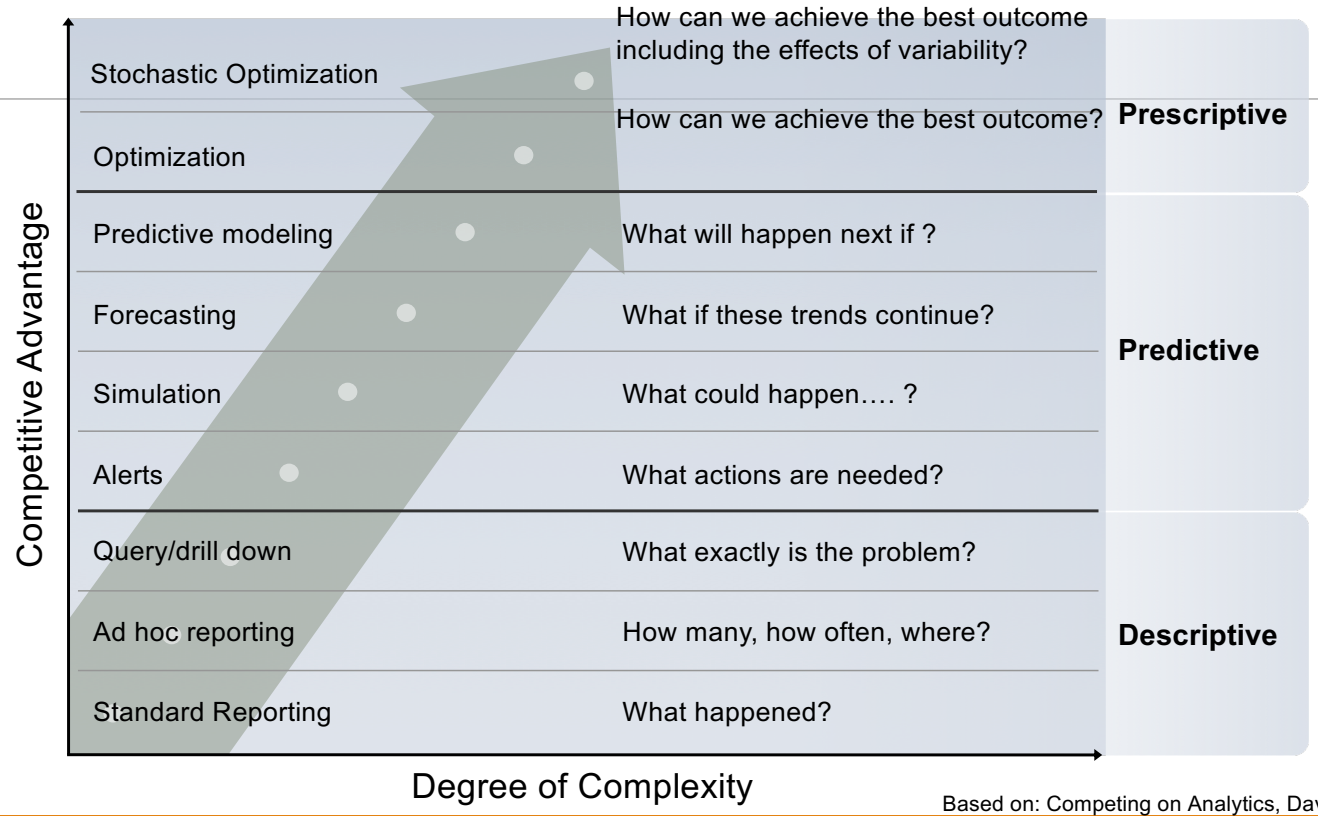


Data sources:
Business automation
Instrumentation
Sensors
Web 2.0
Expert knowledge
"real world physics"

Model:
a mathematical or algorithmic representation of reality intended to explain or predict some aspect of it

Decision executed automatically or by people

Analytics Landscape



Based on: Competing on Analytics, Davenport and Harris, 2007

Example: Courses for a Student

- Decision: Student deciding which courses to take for their program
- Data
  - About courses
  - About faculties
  - About job opportunities
  - About research opportunities and industry trends
  - Private: what the student wants to do
- Analysis
  - **Descriptive**: Courses offered in different semesters; Teachers offering courses
  - **Predictive**: How fill will be a particular class next semester?
  - **Prescriptive**: Should a student a particular course?

Example: Health During a Pandemic

- Decision: Individual staying healthy during a pandemic like COVID19
- Data
  - About cases
  - About mitigation steps: e.g., mask wearing restrictions and practices, lockdowns, hospital conditions
  - Private: pre-existing health conditions
- Analysis
  - **Descriptive**: Regions with high and low cases
  - **Predictive**: Does wearing mask help reduce cases?
  - **Prescriptive**:
    - Whether to eat inside a restaurant?
    - How to make an urgent road trip ?
    - How to hold classes at a University?

Resources: https://github.com/biplav-s/covid19-info/wiki/Important-Information-About-COVID19

Sub-areas of AI

- **Representation**: formal representation of knowledge.
  - Illustration: entities and their relationships, like last Russian Czar's family tree
  - Methods: Ontology, knowledge graph, word embedding, "Model"
- **Reasoning**: deriving conclusions from formally represented knowledge.
  - Illustration: Modus ponen – P implies Q. P is True. Hence Q must be true.
  - Methods: Deduction, Induction, Abduction, Proposition logic, First-order logic, Fuzzy logic
- **Learning**: drawing insights from data
  - Illustration: predict COVID cases in USA by end of the month
  - Methods: Machine Learning – Classification, Clustering, Association; Deep Neural Network

- *Additionally*, ***human interaction considerations***:
  - Collaborative assistants
  - Explanations

Data – The Fuel for AI

Types of Data -1

- Structured Data

- Open Data is one source
  - Often easiest to get but with issues (e.g., at aggregate level, with gaps, imprecise semantics)

- Social is another promising data
  - People are anyway generating it (People-as-sensors)
  - However, social sites have varying data reuse permissions, license costs, access limits
  - Big data techniques already being used here

Types of Data -2

- Use sensor data if available
  - Internet of Things (IoT) and big data techniques are relevant
  - Most prevalent in health, environment and transportation
- Key is to release the fused data also for reuse

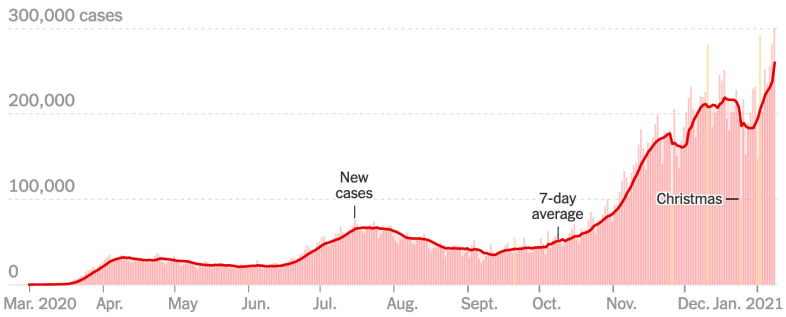
Case Study: COVID-19

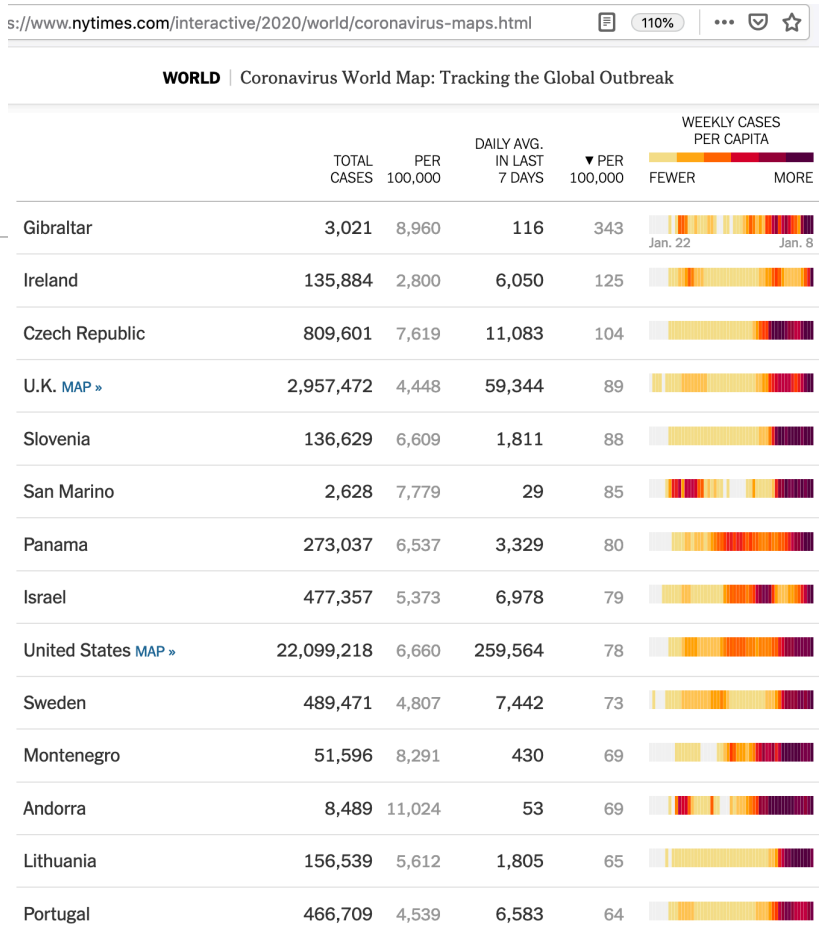
https://www.nytimes.com/interactive/2020/world/coronavirus-maps.html

Q: How good has a country done?

Updated January 9, 2021, 2:10 P.M. E.T.

Leer en español



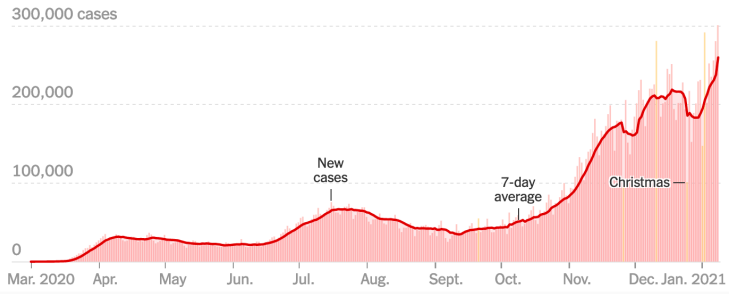


Case Study: COVID-19

https://www.nytimes.com/interactive/2020/us/coronavirus-us-cases.html

Q: How good has a US state done?

Updated January 9, 2021, 2:10 P.M. E.T.
Leer en español



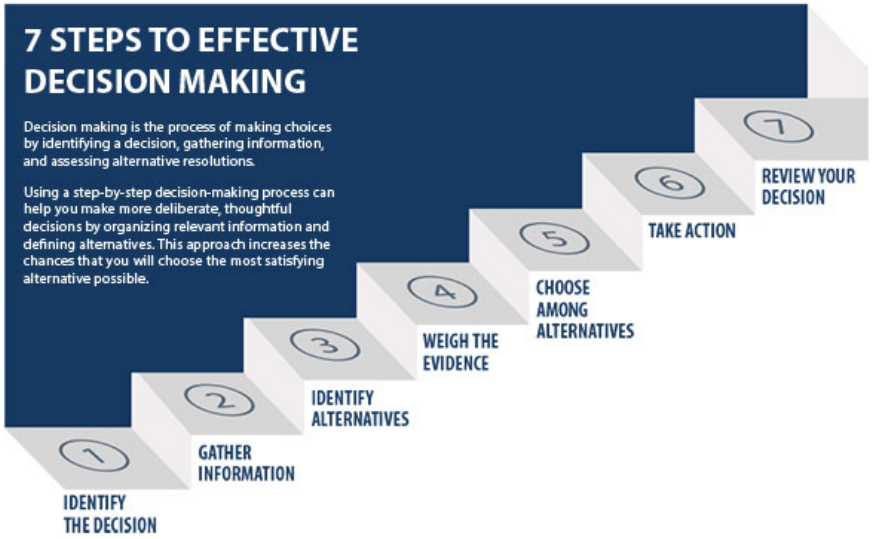
s://www.nytimes.com/interactive/2020/us/coronavirus-us-cases.html 110%

U.S. | Coronavirus in the U.S.: Latest Map and Case Count

	TOTAL CASES	PER 100,000	DAILY AVG. IN LAST 7 DAYS	▼ PER 100,000	WEEKLY CASES PER CAPITA (FEWER – MORE; March 1 – Jan. 8)
+ Arizona MAP »	611,900	8,407	10,268	141	
+ Rhode Island MAP »	97,614	9,214	1,381	130	
+ Tennessee MAP »	629,047	9,211	7,885	115	
+ Utah MAP »	301,110	9,392	3,500	109	
+ Oklahoma MAP »	324,875	8,210	4,236	107	
+ California MAP »	2,658,333	6,728	42,013	106	
+ South Carolina MAP »	350,084	6,799	5,238	102	
+ Kentucky MAP »	300,379	6,723	4,287	96	
+ Massachusetts MAP »	419,721	6,090	6,363	92	
+ Arkansas MAP »	248,860	8,246	2,774	92	
+ Florida MAP »	1,449,244	6,748	17,991	84	
+ North Carolina MAP »	605,187	5,770	8,744	83	
+ Alabama MAP »	399,150	8,141	4,077	83	
+ Connecticut MAP »	205,994	5,778	2,898	81	
+ Louisiana MAP »	341,431	7,345	3,737	80	

Decisions – AI as a Decision-Support Tool

The Quality of Everyday Decisions



Major variability due to:

- Emotions
- Biases
- Increasing data volume
- Cognitive ability to process
  - Decreases under stress and constraints
  - Decreases with age*

Source: https://www.umassd.edu/fycm/decision-making/process/

*** Source**: A Review of Decision-Making Processes: Weighing the Risks and Benefits of Aging, Mara Mather, https://www.ncbi.nlm.nih.gov/books/NBK83778/

Evidence #1:
Poor Medical Adherence

Taking medicines

- 20 -30 % of medication prescriptions are never filled
- ~50 % of medications for chronic disease are not taken as prescribed

Impact

- causes 125,000 deaths, at least 10 percent of hospitalizations
- Costs the American health care system between $100 billion and $289 billion a year.

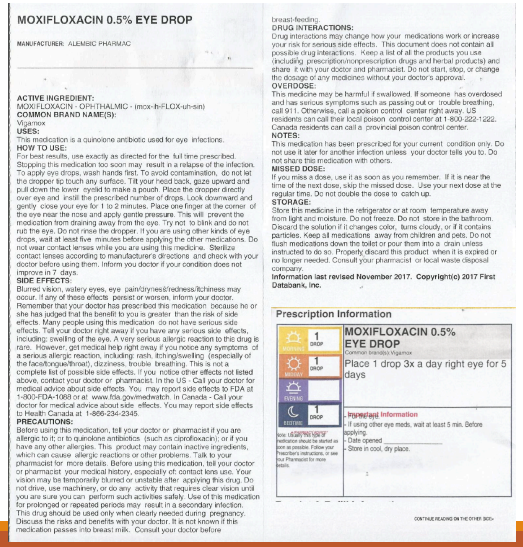
Finding relevant guidance is hard, one reason for non-adherence and high costs in health

Sources:

- Medication Nonadherence, A Diagnosable and Treatable Medical Condition, Zachary A. Marcum, Mary Ann Sevick, Steven M. Handler, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3976600/, 2013.
- https://www.nytimes.com/2017/04/17/well/the-cost-of-not-taking-your-medicine.html

Example:

Hard to understand medicine's information



MOXIFLOXACIN 0.5% EYE DROP

MANUFACTURER: ALEMIC PHARMAC

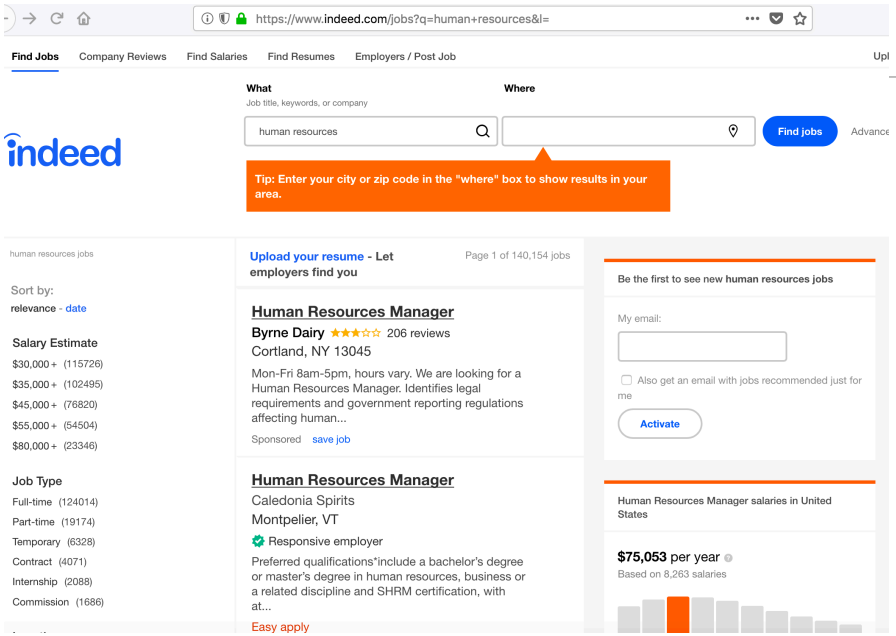
ACTIVE INGREDIENT:
MOXIFLOXACIN - OPHTHALMIC - (mox-ih-FLOX-uh-sin)
COMMON BRAND NAME(S):
Vigamox
USES:
This medication is a quinolone antibiotic used for eye infections.
HOW TO USE:
For best results, use exactly as directed for the full time prescribed. Stopping this medication too soon may result in a relapse of the infection. To apply eye drops, wash hands first. To avoid contamination, do not let the dropper tip touch any surface. Tilt your head back, gaze upward and pull down the lower eyelid to make a pouch. Place the dropper directly over eye and instill the prescribed number of drops. Look downward and gently close your eye for 1 to 2 minutes. Place one finger at the corner of the eye near the nose and apply gentle pressure. This will prevent the medication from draining away from the eye. Try not to blink and do not rub the eye. Do not rinse the dropper. If you are using other kinds of eye drops, wait at least five minutes before applying the other medications. Do not wear contact lenses while you are using this medicine. Sterilize contact lenses according to manufacturer's directions and check with your doctor before using them. Inform you doctor if your condition does not improve in 7 days.
SIDE EFFECTS:
Blurred vision, watery eyes, eye pain/dryness/redness/itchiness may occur. If any of these effects persist or worsen, inform your doctor. Remember that your doctor has prescribed this medication because he or she has judged that the benefit to you is greater than the risk of side effects. Many people using this medication do not have serious side effects. Tell your doctor right away if you have any serious side effects, including: swelling of the eye. A very serious allergic reaction to this drug is rare. However, get medical help right away if you notice any symptoms of a serious allergic reaction, including: rash, itching/swelling (especially of the face/tongue/throat), dizziness, trouble breathing. This is not a complete list of possible side effects. If you notice other effects not listed above, contact your doctor or pharmacist. In the US - Call your doctor for medical advice about side effects. You may report side effects to FDA at 1-800-FDA-1088 or at www.fda.gov/medwatch. In Canada - Call your doctor for medical advice about side effects. You may report side effects to Health Canada at 1-866-234-2345.
PRECAUTIONS:
Before using this medication, tell your doctor or pharmacist if you are allergic to it; or to quinolone antibiotics (such as ciprofloxacin); or if you have any other allergies. This product may contain inactive ingredients, which can cause allergic reactions or other problems. Talk to your pharmacist for more details. Before using this medication, tell your doctor or pharmacist your medical history, especially of: contact lens use. Your vision may be temporarily blurred or unstable after applying this drug. Do not drive, use machinery, or do any activity that requires clear vision until you are sure you can perform such activities safely. Use of this medication for prolonged or repeated periods may result in a secondary infection. This drug should be used only when clearly needed during pregnancy. Discuss the risks and benefits with your doctor. It is not known if this medication passes into breast milk. Consult your doctor before breast-feeding.
DRUG INTERACTIONS:
Drug interactions may change how your medications work or increase your risk for serious side effects. This document does not contain all possible drug interactions. Keep a list of all the products you use (including prescription/nonprescription drugs and herbal products) and share it with your doctor and pharmacist. Do not start, stop, or change the dosage of any medicines without your doctor's approval.
OVERDOSE:
This medicine may be harmful if swallowed. If someone has overdosed and has serious symptoms such as passing out or trouble breathing, call 911. Otherwise, call a poison control center right away. US residents can call their local poison control center at 1-800-222-1222. Canada residents can call a provincial poison control center.
NOTES:
This medication has been prescribed for your current condition only. Do not use it later for another infection unless your doctor tells you to. Do not share this medication with others.
MISSED DOSE:
If you miss a dose, use it as soon as you remember. If it is near the time of the next dose, skip the missed dose. Use your next dose at the regular time. Do not double the dose to catch up.
STORAGE:
Store this medicine in the refrigerator or at room temperature away from light and moisture. Do not freeze. Do not store in the bathroom. Discard the solution if it changes color, turns cloudy, or if it contains particles. Keep all medications away from children and pets. Do not flush medications down the toilet or pour them into a drain unless instructed to do so. Properly discard this product when it is expired or no longer needed. Consult your pharmacist or local waste disposal company.
Information last revised November 2017. Copyright(c) 2017 First Databank, Inc.

Prescription Information

MOXIFLOXACIN 0.5% EYE DROP

Place 1 drop 3x a day right eye for 5 days

Evidence #2: Matching Demand to Supply of Jobs is Inadequate Demand-Supply Gap in Jobs Market [1] and Yet, Low Work Satisfaction/ Engagement [2]



Job search at a portal

- Finding jobs was generally hard around the world (Dec 2019), except for in tight labor markets like US (3.5% unemployment)
- Workforce satisfaction/ engagement was generally low around the world – people did not find jobs they were match for [1,2]
- COVID-19 impact [3]:
  - *Nearly half of global workforce at risk of losing livelihoods in informal sector*
  - *9-12% job loss in the formal sector around the world*
  - *14.7% unemployment in US by end of April 2020 [4]*

1. **Source**: Global Skills Trends, Training Needs and Lifelong Learning Strategies for the Future of Work, ILO & OECD Report 2018, http://www.g20.utoronto.ca/2018/g20_global_skills_trends_and_lll_oecd-ilo.pdf
2. **Source**: For 2016, job satisfaction: US – 32%, Global – 13%, https://www.gallup.com/workplace/236495/worldwide-employee-engagement-crisis.aspx
3. **https://www.ilo.org/global/about-the-ilo/newsroom/news/WCMS_743036/lang--en/index.htm**
4. **https://www.bls.gov/news.release/empsit.nr0.htm**

Decision Imperative: Corona Virus Pandemic

Emerging Scenario Around the World*

- Millions of cases, hundreds of thousands of deaths
- Businesses disrupted, millions going out of business
- Millions loosing jobs

* Numbers changing continuously; see reference for details

Decisions Need to be Made

- About disease
  - Understand disease
  - Tackle disease
- Understand impact to society: economy, supply chain
- Advise on actions to take
  - Individual
  - Group
  - Societal policy

Resource: https://github.com/biplav-s/covid19-info/wiki/Important-Information-About-COVID19

Pressing Issue: Distribution of Vaccines

- Problem: Limited supply, larger demand; How do distribute equitably, fairly and efficiently
- Possible (automated) solutions
  - Random: pick receiver based on random choice
    - **Benefit**: Easy to implement
    - **Problems**: Equitable but not fair, receiver may not be at risk or not want it, others wanting it may not get it
    - **Question**: assumes we can give vaccine quickly to the selected person
  - Prioritized random: make a prioritized list of groups, assign randomly in each group
    - **Benefit**: identifies affected groups
    - **Problems**: receiver may not want the vaccine
    - **Question**: who comes up with groups?, is it rewarding groups who have not been taking precautions ? Assumes we can give vaccine quickly to the selected person
  - ...
  - Benefit-cost: based on contribution to economy
    - **Benefit**: efficient

AI-Based Decision-Support for COVID-19

- Understanding the disease
  - Disease spread and simulation models
  - Insights by visualization
- Tackling the disease
  - Tracking people's movement
  - Fever detection via images
  - Understanding mental depression from social posts
  - Fighting fake news
- Understanding impact
  - Economic – job loss, industrial growth
  - Supply Chain
  - Risks

- Individual actions
  - *Screening/ triage tools*
- Group actions
  - *Models for when to open economy*
  - *Contact tracing*
  - *Matching producers and consumers: food, medical supplies*
- Policy actions
  - *Understanding impact of policy choices (e.g. lockdowns, travel restrictions)*
  - *Design of economic interventions*
- AI Community's Learning
  - *Data sources: Structured, Text – Research papers, Image / Video*
  - *Sharing and reuse of models and data is important*
  - *Lots of hackathons*

Resource: https://github.com/biplav-s/covid19-info/wiki/Important-Information-About-COVID19

Guideline: Human Impact of AI

- We study technology (AI) but it works with data
- Data, when from people or about people, can have issues like bias
  - **Example**: data reveals a view which is influenced by data collection practices
  - **Difference**: World as it is, world according to data and **world as it should be**
- The course and instructor believes in
  - Not promoting bias of any kind
  - Respecting everyone regardless of background

Lecture 1: Concluding Comments

- We did a quick overview of the course
- Looked at AI, Data and Decisions
- Course will focus on
  - Practical methods to derive insights from open data
  - Evaluation will be by via project, paper and quizzes
  - **Bring your ideas to your project**
- Exciting techniques to learn to impact the world around us

Concluding Segment

About Next Lecture – Lecture 2

Lecture 2: Data

- Types
  - Structured,
  - Semi-structured,
  - Unstructured,
- By media: text, audio, video, multi-media;
- Open Data