

# **SWS3023**

# **WEB MINING**

**INTRODUCTION**

# ABOUT ME



**LEK Hsiang Hui**

**Senior Lecturer**

**Department of Information Systems  
and Analytics**

**Teaches Courses on:**

- Analytics
- Software Engineering

**Undergraduates and Executives**

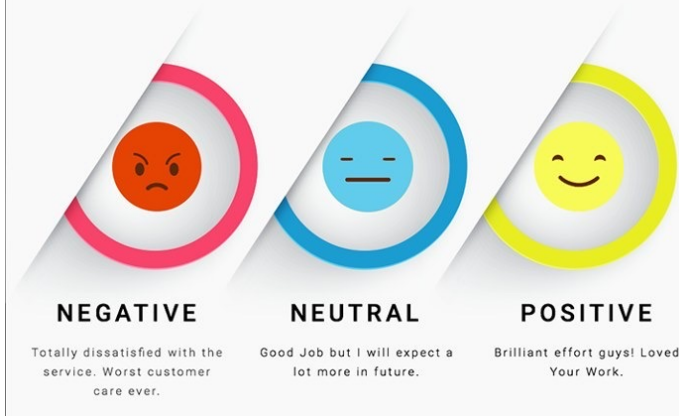
# ABOUT ME



**Did PhD in the area Natural Language Processing (Sentiment Analysis)**

# ABOUT ME

## SENTIMENT ANALYSIS



Scott, Sydney "Just point and shoot" [More like this](#)



+

Picture quality is brilliant. It's small, light and does everything you'd want - and more!

-

Power button is rather difficult to access.

Was this review helpful? [Yes](#) [No](#)

1 of 1 people found this review helpful



**Picture Quality**

brilliant



**Size**



small

**Weight**

light



**Power Button**



difficult to access

# about us

Based in Singapore's silicon valley, Blk 73 Launchpad, we are a big data analytics company, revolutionizing global and local brand analytics ranking and profiling. With billions of online conversations, we have the data and technologies to determine any brands' relative brand position globally. We have the ability to automatically conduct extensive user profiling for brands wanting to know more about their customers' preferences and desires. Through workshops and trainings, we help brands make sense of data and get them started with digital marketing.

# what we offer

## CUSTOMER INSIGHTS

Want to know what netizens are talking about your brand, competitors or a brand? How can you improve your services or business offerings?

## INDUSTRY KNOWLEDGE

What is the latest trend happening - for your brand, your industry or target audience? What can we learn from others in your industry to help you in your marketing strategy?

## INFLUENCER PROFILING

Who are your influencers? Can you turn them into leads? Or use them to bring in leads?

Data  
made  
simple



## REPORTS

With our technology and the immense data that are available on the web, we can provide customer insights, industry knowledge and influencer profiling.



## TRAININGS

We conduct trainings and workshops to help brands solve problems using their data and to provide knowledge to propel your marketing strategies. Let us help you understand analytics and marketing tactics.



## TALK TO US

It's just a day to find out more about the services that we offer. We'll be glad to help you make life simple just over coffee and tea. You'd just need to let us do the thinking!

email: [enquiry@standices.com](mailto:enquiry@standices.com)

Also Co-founded a  
Big Data Analytics  
Company

# CONTACT



<https://www.facebook.com/hsianghui>



[@hsianghui](#)



<https://www.linkedin.com/in/hsianghui/>



[hsianghui@nus.edu.sg](mailto:hsianghui@nus.edu.sg)

# TEACHING ASSISTANT



Mr TAN Qiu Yu

✉ [qiuyu.87@gmail.com](mailto:qiuyu.87@gmail.com)

# OBJECTIVES AND OVERVIEW

## Aim:

- Provide a good understanding of predictive analytics and its applications
- Provide a systematic approach to mine web content

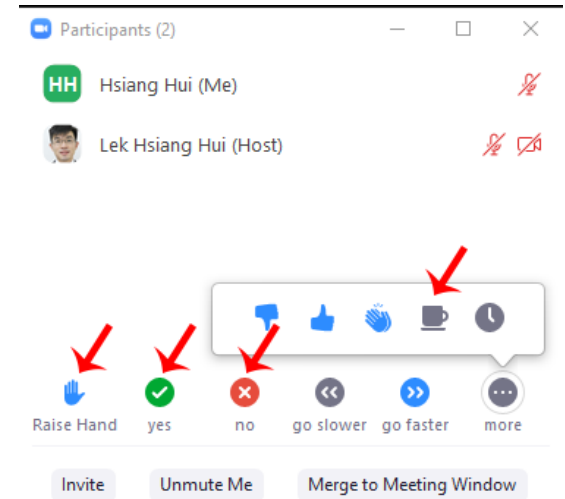
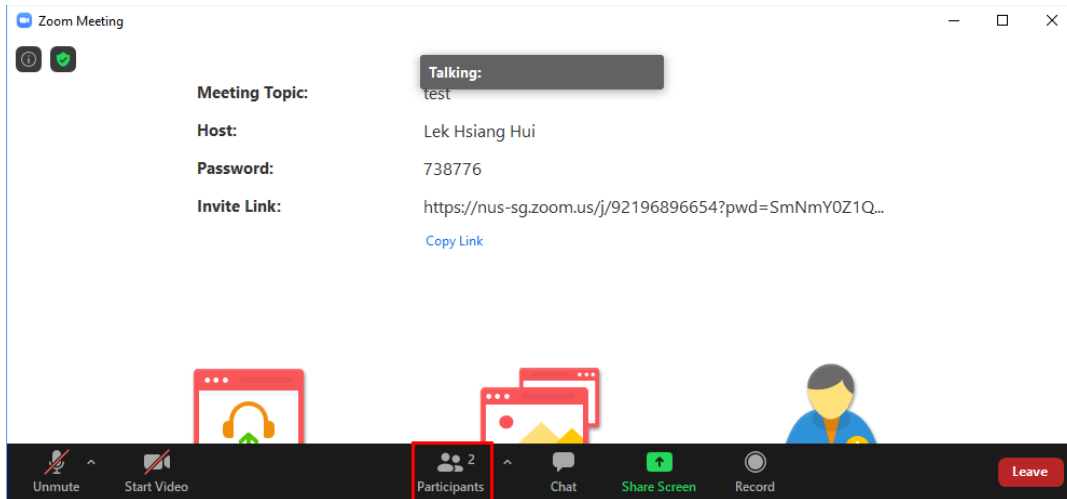
## Objectives

- Understand the conceptual foundations of predictive analytics
- Be able to programmatically mine web content
- Apply predictive modeling on web data



# SOME HOUSE RULES (ZOOM)

Please **respond** in Zoom



# TOPICS

**Introduction to Analytics  
and Web Mining (29 May)**

**Predictive Analytics I**

**Predictive Analytics II**

**Mining Web Content I**

**Mining Web Content II**

**Mining Web Content III**

**Recommender Systems**



Lectures include  
interactive in-class  
activities

# TEACHING/CONSULTATION SCHEDULE\*

July 2022					
Mon	Tues	Wed	Thurs	Fri	Sat
<b>11</b> <b>CRISP-DM &amp; Predictive Analytics I</b> (10-12pm) <b>Lab 1</b> (1-3pm)  <b>1<sup>st</sup> consultation</b> (3-6pm)	<b>12</b> <b>Predictive Analytics II</b> (10-12pm)  <b>Lab 2</b> (1-3pm)  <b>1<sup>st</sup> consultation</b> (3-6pm)	<b>13</b> <b>Mining Web Content I</b> (10-12pm)  <b>Lab 3</b> (1-3pm)  <b>1<sup>st</sup> consultation</b> (3-6pm)	<b>14</b> <b>Mining Web Content II</b> (10-12pm)  <b>Lab 4</b> (1-3pm)  <b>2<sup>nd</sup> consultation</b> (3-6pm)	<b>15</b> <b>Mining Web Content III</b> (10-12pm)  <b>Lab 5</b> (1-3pm)  <b>2<sup>nd</sup> consultation</b> (3-6pm)	<b>16</b> <b>Ad hoc consultation/ help from TA</b> (10-6pm)
<b>18</b> <b>Recommender System</b> (10am-12pm)  <b>Lab 6</b> (1-3pm)  <b>2<sup>nd</sup> consultation</b> (3pm-6pm)	<b>19</b> <b>3<sup>rd</sup> consultation</b> (10-6pm)	<b>20</b> <b>Ad hoc consultation/ help from TA</b> (10-6pm)	<b>21</b> <b>4<sup>th</sup> consultation</b>  <b>Ad hoc help from TA</b> (10am-6pm)	<b>22</b> <b>4<sup>th</sup> consultation</b>  <b>Ad hoc help from TA</b> (10am-6pm)	<b>23</b> <b>Ad hoc consultation</b> (10-6pm)
<b>25</b> <b>Ad hoc consultation/ help from TA</b> (10-6pm)	<b>26</b> <b>Ad hoc consultation/ help from TA</b> (10-6pm)	<b>27</b> <b>Poster Submission</b>	<b>28</b> <b>Project Showcase</b> (12-6pm)	<b>29</b>	<b>30</b>

# TEACHING/CONSULTATION SCHEDULE\*

July 2022

Mon	Tues	Wed	Thurs	Fri	Sat
<b>11</b> <b>CRISP-DM &amp; Predictive Analytics I</b> (10-12pm) <b>Lab 1</b> (1-3pm) <b>1<sup>st</sup> consultation</b> (3-6pm)	<b>12</b> <b>Predictive Analytics II</b> (10-12pm) <b>Lab 2</b> (1-3pm) <b>1<sup>st</sup> consultation</b> (3-6pm)	<b>13</b> <b>Mining Web Content I</b> (10-12pm) <b>Lab 3</b> (1-3pm) <b>1<sup>st</sup> consultation</b> (3-6pm)	<b>14</b> <b>Mining Web Content II</b> (10-12pm) <b>Lab 4</b> (1-3pm) <b>2<sup>nd</sup> consultation</b> (3-6pm)	<b>15</b> <b>Mining Web Content III</b> (10-12pm) <b>Lab 5</b> (1-3pm) <b>2<sup>nd</sup> consultation</b> (3-6pm)	<b>16</b> <b>Ad hoc consultation/ help from TA</b> (10-6pm)
<b>18</b> <b>Recommender System</b> (10am-12pm) <b>Lab 6</b> (1-3pm) <b>2<sup>nd</sup> consultation</b> (3pm-6pm)	<b>19</b> <b>3<sup>rd</sup> consultation</b> (10-6pm)	<b>20</b> <b>Ad hoc consultation/ help from TA</b> (10-6pm)	<b>21</b> <b>4<sup>th</sup> consultation</b> <b>Ad hoc help from TA</b> (10am-6pm)	<b>22</b> <b>4<sup>th</sup> consultation</b> <b>Ad hoc help from TA</b> (10am-6pm)	<b>23</b> <b>Ad hoc consultation</b> (10-6pm)
<b>25</b> <b>Ad hoc consultation/ help from TA</b> (10-6pm)	<b>26</b> <b>Ad hoc consultation/ help from TA</b> (10-6pm)	<b>27</b> <b>Poster Submission</b>	<div>                     3 additional lectures that provide the classic techniques to perform predictive analytics and/or handle web data (in red)                 </div>		

# TEACHING/CONSULTATION SCHEDULE\*

July 2022

Mon	Tues	Wed	Thurs	Fri	Sat
<b>11</b> <b>CRISP-DM &amp; Predictive Analytics I</b> (10-12pm) <b>Lab 1</b> (1-3pm) <b>1<sup>st</sup> consultation</b> (3-6pm)	<b>12</b> <b>Predictive Analytics II</b> (10-12pm) <b>Lab 2</b> (1-3pm) <b>1<sup>st</sup> consultation</b> (3-6pm)	<b>13</b> <b>Mining Web Content I</b> (10-12pm) <b>Lab 3</b> (1-3pm) <b>1<sup>st</sup> consultation</b> (3-6pm)	<b>14</b> <b>Mining Web Content II</b> (10-12pm) <b>Lab 4</b> (1-3pm) <b>2<sup>nd</sup> consultation</b> (3-6pm)	<b>15</b> <b>Mining Web Content III</b> (10-12pm) <b>Lab 5</b> (1-3pm) <b>2<sup>nd</sup> consultation</b> (3-6pm)	<b>16</b> <b>Ad hoc consultation/ help from TA</b> (10-6pm)
<b>18</b> <b>Recommender System</b> (10am-12pm) <b>Lab 6</b> (1-3pm) <b>2<sup>nd</sup> consultation</b> (3pm-6pm)	<b>19</b> <b>3<sup>rd</sup> consultation</b> (10-6pm)	<b>20</b> <b>Ad hoc consultation/ help from TA</b> (10-6pm)	<b>21</b> <b>4<sup>th</sup> consultation</b> <b>Ad hoc help from TA</b> (10am-6pm)	<b>22</b> <b>4<sup>th</sup> consultation</b> <b>Ad hoc help from TA</b> (10am-6pm)	<b>23</b> <b>Ad hoc consultation</b> (10-6pm)
<b>25</b> <b>Ad hoc consultation/ help from TA</b> (10-6pm)	<b>26</b> <b>Ad hoc consultation/ help from TA</b> (10-6pm)	<b>27</b> <b>Poster Submission</b>	3 additional lectures that teach you how to mine pretty much any website (in red)		

# TEACHING/CONSULTATION SCHEDULE\*

July 2022

Mon	Tues	Wed	Thurs	Fri	Sat
<b>11</b> <b>CRISP-DM &amp; Predictive Analytics I</b> (10-12pm) <b>Lab 1</b> (1-3pm) <b>1<sup>st</sup> consultation</b> (3-6pm)	<b>12</b> <b>Predictive Analytics II</b> (10-12pm) <b>Lab 2</b> (1-3pm) <b>1<sup>st</sup> consultation</b> (3-6pm)	<b>13</b> <b>Mining Web Content I</b> (10-12pm) <b>Lab 3</b> (1-3pm) <b>1<sup>st</sup> consultation</b> (3-6pm)	<b>14</b> <b>Mining Web Content II</b> (10-12pm) <b>Lab 4</b> (1-3pm) <b>2<sup>nd</sup> consultation</b> (3-6pm)	<b>15</b> <b>Mining Web Content III</b> (10-12pm) <b>Lab 5</b> (1-3pm) <b>2<sup>nd</sup> consultation</b> (3-6pm)	<b>16</b> <b>Ad hoc consultation/ help from TA</b> (10-6pm)
<b>18</b> <b>Recommender System</b> (10am-12pm) <b>Lab 6</b> (1-3pm) <b>2<sup>nd</sup> consultation</b> (3pm-6pm)	<b>19</b> <b>3<sup>rd</sup> consultation</b> (10-6pm)	<b>20</b> <b>Ad hoc consultation/ help from TA</b> (10-6pm)	<b>21</b> <b>4<sup>th</sup> consultation</b> <b>Ad hoc help from TA</b> (10am-6pm)	<b>22</b> <b>4<sup>th</sup> consultation</b> <b>Ad hoc help from TA</b> (10am-6pm)	<b>23</b> <b>Ad hoc consultation</b> (10-6pm)
<b>25</b> <b>Ad hoc consultation/ help from TA</b> (10-6pm)	<b>26</b> <b>Ad hoc consultation/ help from TA</b> (10-6pm)	<b>27</b> <b>Pos Sub</b>	<div>                         Lab sessions equip you with hands on skills how to handle data with Python (e.g. Basics of Python, Pandas, etc)                     </div>		<b>30</b>

# TEACHING/CONSULTATION SCHEDULE\*

July 2022

Mon	Tues	Wed	Thurs	Fri	Sat
<b>11</b> <b>CRISP-DM &amp; Predictive Analytics I</b> (10-12pm) <b>Lab 1</b> (1-3pm) <b>1<sup>st</sup> consultation</b> (3-6pm)	<b>12</b> <b>Predictive Analytics II</b> (10-12pm) <b>Lab 2</b> (1-3pm) <b>1<sup>st</sup> consultation</b> (3-6pm)	<b>13</b> <b>Mining Web Content I</b> (10-12pm) <b>Lab 3</b> (1-3pm) <b>1<sup>st</sup> consultation</b> (3-6pm)	<b>14</b> <b>Mining Web Content II</b> (10-12pm) <b>Lab 4</b> (1-3pm) <b>2<sup>nd</sup> consultation</b> (3-6pm)	<b>15</b> <b>Mining Web Content III</b> (10-12pm) <b>Lab 5</b> (1-3pm) <b>2<sup>nd</sup> consultation</b> (3-6pm)	<b>16</b> <b>Ad hoc consultation/ help from TA</b> (10-6pm)
<b>18</b> <b>Recommender System</b> (10am-12pm) <b>Lab 6</b> (1-3pm) <b>2<sup>nd</sup> consultation</b> (3pm-6pm)	<b>19</b> <b>3<sup>rd</sup> consultation</b> (10-6pm)	<b>20</b> <b>Ad hoc consultation/ help from TA</b> (10-6pm)	<b>21</b> <b>4<sup>th</sup> consultation</b> <b>Ad hoc help from TA</b> (10am-6pm)	<b>22</b> <b>4<sup>th</sup> consultation</b> <b>Ad hoc help from TA</b> (10am-6pm)	<b>23</b> <b>Ad hoc consultation</b> (10-6pm)
<b>25</b> <b>Ad hoc consultation/ help from TA</b> (10-6pm)	<b>26</b> <b>Ad hoc consultation/ help from TA</b> (10-6pm)	<b>27</b> <b>Poster Submission</b>	<div>                         Lab sessions to learn how to perform modeling with Python                     </div>		

# TEACHING/CONSULTATION SCHEDULE\*

July 2022

Mon	Tues	Wed	Thurs	Fri	Sat
<b>11</b> <b>CRISP-DM &amp; Predictive Analytics I</b> (10-12pm) <b>Lab 1</b> (1-3pm) <b>1<sup>st</sup> consultation</b> (3-6pm)	<b>12</b> <b>Predictive Analytics II</b> (10-12pm) <b>Lab 2</b> (1-3pm) <b>1<sup>st</sup> consultation</b> (3-6pm)	<b>13</b> <b>Mining Web Content I</b> (10-12pm) <b>Lab 3</b> (1-3pm) <b>1<sup>st</sup> consultation</b> (3-6pm)	<b>14</b> <b>Mining Web Content II</b> (10-12pm) <b>Lab 4</b> (1-3pm) <b>2<sup>nd</sup> consultation</b> (3-6pm)	<b>15</b> <b>Mining Web Content III</b> (10-12pm) <b>Lab 5</b> (1-3pm) <b>2<sup>nd</sup> consultation</b> (3-6pm)	<b>16</b> <b>Ad hoc consultation/ help from TA</b> (10-6pm)
<b>18</b> <b>Recommender System</b> (10am-12pm) <b>Lab 6</b> (1-3pm) <b>2<sup>nd</sup> consultation</b> (3pm-6pm)	<b>19</b> <b>3<sup>rd</sup> consultation</b> (10-6pm)	<b>20</b> <b>Ad hoc consultation/ help from TA</b> (10-6pm)	<b>21</b> <b>4<sup>th</sup> consultation</b> <b>Ad hoc help from TA</b> (10am-6pm)	<b>22</b> <b>4<sup>th</sup> consultation</b> <b>Ad hoc help from TA</b> (10am-6pm)	<b>23</b> <b>Ad hoc consultation</b> (10-6pm)
<b>25</b> <b>Ad hoc consultation/ help from TA</b> (10-6pm)	<b>26</b> <b>Ad hoc consultation/ help from TA</b> (10-6pm)	<b>27</b> <b>Poster Submission</b>	<div>                     Lab sessions to practice web scraping                 </div>		



# TEACHING/CONSULTATION SCHEDULE\*

July 2022					
Mon	Tues	Wed	Thurs	Fri	Sat
<b>11</b> <b>CRISP-DM &amp; Predictive Analytics I</b> (10-12pm) <b>Lab 1</b> (1-3pm)  <b>1<sup>st</sup> consultation</b> (3-6pm)	<b>12</b> <b>Predictive Analytics II</b> (10-12pm)  <b>Lab 2</b> (1-3pm)  <b>1<sup>st</sup> consultation</b> (3-6pm)	<b>13</b> <b>Mining Web Content I</b> (10-12pm)  <b>Lab 3</b> (1-3pm)  <b>1<sup>st</sup> consultation</b> (3-6pm)	<b>14</b> <b>Mining Web Content II</b> (10-12pm)  <b>Lab 4</b> (1-3pm)  <b>2<sup>nd</sup> consultation</b> (3-6pm)	<b>15</b> <b>Mining Web Content III</b> (10-12pm)  <b>Lab 5</b> (1-3pm)  <b>2<sup>nd</sup> consultation</b> (3-6pm)	<b>16</b> <b>Ad hoc consultation/ help from TA</b> (10-6pm)
<b>18</b> <b>Recommender System</b> (10am-12pm)  <b>Lab 6</b> (1-3pm)  <b>2<sup>nd</sup> consultation</b> (3pm-6pm)	<b>19</b> <b>3<sup>rd</sup> consultation</b> (10-6pm)	<b>20</b> <b>Ad hoc consultation/ help from TA</b> (10-6pm)	<b>21</b> <b>4<sup>th</sup> consultation</b>  <b>Ad hoc help from TA</b> (10am-6pm)	<b>22</b> <b>4<sup>th</sup> consultation</b>  <b>Ad hoc help from TA</b> (10am-6pm)	<b>23</b> <b>Ad hoc consultation</b> (10-6pm)
<b>25</b> <b>Ad hoc consultation/ help from TA</b> (10-6pm)	<b>26</b> <b>Ad hoc consultation/ help from TA</b> (10-6pm)	<b>27</b> <b>P</b> <b>S</b>	<div>           Lab session on            Recommender Systems         </div>		<b>30</b>

# GROUP PROJECT

**Work in groups of 4** (tentatively)

**Self-propose an interesting analytics problem that requires some form of predictive analytics using web data**

- Formulate a strategy to mine the web data
- Use appropriate analytics techniques to process the data
- Guidance will be provided along the way
- Planned consultations with the lecturer, ad-hoc consultations with the lecturer/TA

**4 planned consultations with lecturer**

- 1<sup>st</sup> consultation – Problem formation and project scoping
- 2<sup>nd</sup> consultation – Formulate strategies to mine sites (web scraping)
- 3<sup>rd</sup> consultation – Project fine-tuning (analysis)
- 4<sup>th</sup> consultation – Project fine-tuning (presentation)

# TEACHING/CONSULTATION SCHEDULE\*

July 2022

Mon	Tues	Wed	Thurs	Fri	Sat
<p>11 <b>CRISP-DM &amp; Predictive Analytics I</b> (10-12pm) <b>Lab 1</b> (1-3pm)</p> <p><b>1<sup>st</sup> consultation</b> (3-6pm)</p>	<p>12 <b>Predictive Analytics II</b> (10-12pm) <b>Lab 2</b> (1-3pm)</p> <p><b>1<sup>st</sup> consultation</b> (3-6pm)</p>	<p>13 <b>Mining Web Content I</b> (10-12pm) <b>Lab 3</b> (1-3pm)</p> <p><b>1<sup>st</sup> consultation</b> (3-6pm)</p>	<p>14 <b>Mining Web Content II</b> (10-12pm) <b>Lab 4</b> (1-3pm)</p> <p><b>2<sup>nd</sup> consultation</b> (3-6pm)</p>	<p>15 <b>Mining Web Content III</b> (10-12pm) <b>Lab 5</b> (1-3pm)</p> <p><b>2<sup>nd</sup> consultation</b> (3-6pm)</p>	<p>16 <b>Ad hoc consultation/ help from TA</b> (10-6pm)</p>
<p>18 <b>Recommender System</b> (10am-12pm) <b>Lab 6</b> (1-3pm)</p> <p><b>2<sup>nd</sup> consultation</b> (3pm-6pm)</p>	<p>19 <b>3<sup>rd</sup> consultation</b> (10-6pm)</p>	<p>20 <b>Ad hoc consultation/ help from TA</b> (10-6pm)</p>	<p>21 <b>4<sup>th</sup> consultation</b> <b>Ad hoc help from TA</b> (10am-6pm)</p>	<p>22 <b>4<sup>th</sup> consultation</b> <b>Ad hoc help from TA</b> (10am-6pm)</p>	<p>23 <b>Ad hoc consultation</b> (10-6pm)</p>
<p>25 <b>Ad hoc consultation/ help from TA</b> (10-6pm)</p>	<p>26 <b>Ad hoc consultation/ help from TA</b> (10-6pm)</p>	<p>27 <b>Poster Submission</b></p>	<p>28 <b>Project Showcase</b> (12-6pm)</p>	<p>29</p>	<p>30</p>

# ASSESSMENT\*

## Mainly project

- Participation (in labs/consultation, individual): 10%
- Analysis and Results: 60%
- Project presentation: 30%