Tinashe Michael Tapera, MSc

832 S. 49th Street ● Philadelphia, PA 19143 ● 267-441-7206
Email: tinashemtapera@gmail.com
Website: tinashemtapera.com

SUMMARY

Data Scientist with a focus on quantitative data analysis and computational techniques for psychology and neuroscience research. My mission is to study human behaviour using advanced statistics and machine learning applied to data collected from wearable and mobile devices. Proficient R and Python programmer, with significant experience in the domains of neuroimaging, clinical/health psychology, and people analytics/IO psychology.

EDUCATION

Drexel University September 2016 - June 2018
Accelerated Master of Science in Psychology Philadelphia, PA
Behavioral Data Analysis Concentration
Graduate Minor in Computer Science
Graduating GPA: 3.54 (Cum Laude)

Drexel University September 2013 - June 2017
Bachelor of Science in Psychology Philadelphia, PA
Graduating GPA: 3.51 (Cum Laude)

St. John's College 2007 - 2012
International Baccalaureate Diploma Program Harare, Zimbabwe
Higher level: Physics, Business Studies, English
Standard level: Mathematics, Chemistry, French
32 points

HONORS AND AWARDS

Drexel University
Statistical and Mathematical Sciences Institute Undergraduate Travel Award 2016
College of Arts & Sciences Humanities Fellowship in Psychology 2015
Dean's List Multiple
A. J. Drexel Academic Scholarship 2013 - 2018

St. John's College
Athene Cup for Cultural Person of the Year 2012

PROFESSIONAL ORGANIZATIONS

American Statistical Association (ASA) Member
Eastern North American Region of the International Biometrics Society (ENAR) Member
Society of Behavioral Medicine (SBM) Member
Phi Eta Sigma National Honors Society Member
Drexel University Gospel Choir Member

WORK EXPERIENCE

**Penn Lifespan Informatics &
Neuroimaging Center (PennLINC)**
Neuroimaging Data Analyst Sr.
Supervisor: Theodore Satterthwaite, PhD

October 2018 - Present

Philadelphia, PA

- Use programming tools and technologies to curate, analyse, and present brain imaging data.
- Develop data pipelines for ETL (extract-transform-load) of large-scale imaging data sets between data warehouses in Python, R, and Bash.
- Preprocess and analyse neuroimaging data using cutting edge software (fMRIprep, XCPengine, QSIPrep, ASLPrep).
- Develop and actively maintain multiple data curation software packages in Python and R.
- Published first author in *Frontiers in Neuroinformatics*.

Salesforce
Data Science Intern – People Analytics
Supervisor: Ernest Ng, PhD

June 2017 - September 2017
San Francisco, CA

- Focused on discovery of organizational insight using internal human resources data sets.
- Developed a semi-supervised learning algorithm for tracking employee performance by matching topic models in ongoing feedback data and goal-setting data.
- Designed data processing pipelines for analyses in Python, Bash, and MySQL.
- Examined employee feedback patterns using various text mining techniques:
  o Parts-of-speech tagging,
  o Topic modelling,
  o Sentiment analysis.
- Investigated comorbidity of employees' insurance claims data to dynamically classify claim types and profile employees.
- Carried out ad-hoc statistical investigations as requested.

Arzoo LLC
Private Equity Analyst
Supervisors: Ahmed Makani, Megan Strouss-Rooney, MSW

October 2015 - April 2016
Philadelphia, PA

- Scraped various online business databases to create large datasets of business profiles (5000+ cases).
- Developed a data munging pipeline in Excel using advanced functions to cut down data munging time for interns by almost 75%.
- Taught data preparation techniques to other interns using Excel templates and basic shell scripting.
- Analyzed profiles using financial criteria and seller profile criteria to identify business owners with high potential to sell.
- Initiated negotiations with business owners by email, mail and by phone.

RESEARCH EXPERIENCE

Quantitative Psychology & Statistics Laboratory   2015 - 2019
Graduate Research Assistant   Drexel University
Supervisor: Fengqing Zoe Zhang, PhD

- Investigated advanced data mining and machine learning models for data-intensive research in psychology, neuropsychology, and epidemiology.

- "Momentary Changes in Heart Rate Variability Can Detect Risk for Emotional Eating Episodes."
  - Aimed to predict emotional eating episodes using heart rate data sourced from smartwatches in an emotional eating study.
  - Methods include continuous Heart Rate Variability analysis in the time and frequency domain, feature engineering and selection, and binary classification using support vector machines
  - Achieved moderately high accuracy (77%), sensitivity (79%) and specificity (75%) when predicting the likelihood of an emotional eating episode occurring within the following 30 minute window, based on fluctuations of heart rate.
  - Models can be used to continuously monitor patients at risk of emotional eating episodes, and intervene on positive signals. Manuscript published in *Appetite* (2020).

- "Application of Advanced Data Mining Models to Identify Dietary Patterns Associated with Risk of Cardiovascular Disease."
  - Data sourced from National Health and Nutrition Examination Survey.
  - Aimed to predict levels of cardiovascular disease biomarkers (e.g. cholesterol) using high-dimensional food and behaviour survey responses.
  - Employed and compared principal components analysis + linear regression, and LASSO regression.
  - Achieved adjusted-R^2 of over 0.85 in predicting individual biomarker levels using LASSO approach.
  - Data published in Master's Thesis.

- "Improved Modelling of Smartphone-based Ecological Momentary Assessment Data for Dietary Lapse Prediction."
  - Ecological Momentary Assessment data sourced via periodic mobile app survey.
  - Aimed to predict when participants in dietary adherence study would lapse (i.e. cheat) on their diet by classifying survey responses into lapse or non-lapse probabilities.
  - Developed a novel algorithm for classification in the case of highly imbalanced classes:
    - Hypothesized a neighborhood-based approach to create small subsets of training data based on similarity to the test set observation.
    - Implemented individual classifiers for each subset and test observation employed ensemble rule evaluation for the resulting predictions.
    - Results show moderate improvements in specificity and sensitivity over existing data-balancing and ensemble rule techniques.

- "Identifying Autism Diagnostic Interview: Revised Algorithm Items that Significantly Distinguish Autism Spectrum Disorder and Down Syndrome."

- Data gathered through Autism Diagnostic Review questionnaire (ADI-R), including participants diagnosed with Autism Spectrum Disorder, Down Syndrome, or both.
- Aimed to infer critical diagnostic criteria for early detection and differentiation of diagnoses.
- Built both dichotomous and multiclass classification models for prediction of diagnosis, employing methods such as logistic regression, LASSO, and classification & regression trees such as Random Forest.

- "Modelling Zero-Inflated Distributions in Bouted Physical Activity Data."
  - Data sourced from individual fitness trackers to measure amounts of moderate-to-vigorous physical activity.
  - Assumed zero-inflated compound Tweedie Poisson distribution would best model the change in trajectory of participants' moderate-to-vigorous physical activity (MVPA) bouts over time.
  - Achieved MSE of 11.98 using a fixed effects growth curve model to predict MVPA bouts.

SAMSI Interdisciplinary Undergraduate Workshop
Supervisors: Benjamin Risk, BS and Sujit Ghosh PhD

May 2016

NC State University

- Awarded $900 travel award to participate in the Statistical and Mathematical Sciences Institute Interdisciplinary Undergraduate Workshop 2016.
- Worked with a team of multidisciplinary students to predict multiple sclerosis diagnosis.
- Investigated statistical modelling of Diffusion Tensor Imaging data for MS patients in R and MATLAB.
- Loaded, cleaned, and explored datasets using statistical summaries and visualizations.
- Identified and engineered predictive features for multiple sclerosis diagnosis.
- Successfully modelled MS diagnosis using 3 different logistic regression approaches with 81% classification accuracy.

Laboratory for Innovations in Health-Related Behavior Change
Supervisors: Evan Forman, PhD and Stephanie Goldstein BS

2015

Drexel University

- Data Manager: A Companion Smartphone App to Enhance Dietary Adherence through Predictive Machine Learning.
- Data collection, entry, and cleaning using Excel.
- Managed sensitive participant research and clinical files.
- Conducted in-person and remote assessments for data collection and in-person diagnostic interviewing.
- Independently investigated hypotheses on self-report and baseline measurement discrepancies using SPSS.

PUBLICATIONS

Tapera, T. M., Cieslak, M., Bertolero, M., Adebimpe, A., Aguirre, G. K., Butler, E. R., Cook, P. A., Davila, D., Elliott, M. A., Linguiti, S., Murtha, K., Tackett, W., Detre, J. A., & Satterthwaite, T. D. (2021). FlywheelTools: Data Curation and Manipulation on the Flywheel Platform. *Frontiers in Neuroinformatics*, *15*. https://doi.org/10.3389/fninf.2021.678403

Weinstein, S. M., Vandekar, S. N., Adebimpe, A., **Tapera, T. M.**, Robert-Fitzgerald, T., Gur, R. C., Gur, R. E., Raznahan, A., Satterthwaite, T. D., Alexander-Bloch, A. F., & Shinohara, R. T. (2021). A simple permutation-based test of intermodal correspondence. *Human Brain Mapping*. Published. https://doi.org/10.1002/hbm.25577

Xia, C. H., Barnett, I., **Tapera, T. M.**, Cui, Z., Moore, T., Adebimpe, A., ... & Satterthwaite, T. D. (2021). Mobile Footprinting: Linking Individual Distinctiveness in Mobility Patterns to Mood, Sleep, and Brain Functional Connectivity. *bioRxiv*.

Juarascio, A. S., Crochiere, R. J., **Tapera, T. M.**, Palermo, M., & Zhang, F. (2020). Momentary changes in heart rate variability can detect risk for emotional eating episodes. *Appetite*, *152*, 104698.

Godfrey, M., Hepburn, S., Fidler, D. J., **Tapera, T. M.**, Zhang, F., Rosenberg, C. R., & Lee, N. R. (2019). Autism spectrum disorder (ASD) symptom profiles of children with comorbid Down syndrome (DS) and ASD: A comparison with children with DS-only and ASD-only. *Research in developmental disabilities*, *89*, 83-93

Zhang, F., **Tapera, T. M.**, & Gou, J. (2018). Application of a new dietary pattern analysis method in nutritional epidemiology. *BMC medical research methodology*, *18*(1), 1-10.

CONFERENCE PRESENTATIONS

Valcarel, A. M., **Tapera, T. M.**, Vanderkar, S. N., Adebimpe, A., Roalf, D., Raznahan, A., Satterthwaite, T., Shinohara, R. T., & Linn, K. A. (2019, May). Approaches for Modelling Spatially Varying Associations Between Multimodal Images. Poster presented at Shape Modeling International Conference 2019, Irvine, CA.

Zhang, F. Z., **Tapera, T. M.**, Goldstein. S.P., Forman, E.M. (2018, March). Improved Modeling of Smartphone-based Ecological Momentary Assessment Data for Dietary Lapse Prediction. Talk presented at the ENAR Spring Meeting 2018, Atlanta, GA.

Tapera, T. M., Zhang, F. Z., Forman, E. (2018, February). Novel Neighbourhood-based Classification Algorithms for Dietary Lapse Detection. Talk presented at WELL Center Symposium 2018, Philadelphia, PA.

Godfrey, M., **Tapera, T. M.**, Zhang, F. Z., Lee, N (2018, February). Identifying Autism Diagnostic Interview- Revised Algorithm Items that Significantly Distinguish Autism Spectrum Disorder and Down Syndrome. Talk presented at the International Neuropsychological Society Conference, Washington, D.C.

Tapera, T. M., Ng, E., (2017, September). Using Natural Language Processing to Link Feedback & V2MOM Data. Talk presented at the Tech & Product Intern Demo Day at Salesforce HQ, San Francisco, CA.

Tapera, T. M., Zhang, F. Z., (2017, March). Application of Advanced Data Mining Models to Identify Dietary Patterns Associated with Risk of Disease. Poster presented at the ENAR Spring Meeting 2017, Washington, D.C.

Tapera, T. M., Goldstein, S.P., Evans, B., & Forman, E.M., (2016, April). Does Ecological Momentary Assessment Data Reflect Baseline Self-Report in Weight Loss Treatment? Poster presented at the 37th Annual Meeting and Scientific Sessions of the Society of Behavioral Medicine, Washington, D.C.

Forman, E.M., Goldstein. S.P., Evans, B., Manasse, S.M., Juarascio, A.S., Butryn, M.L., & **Tapera, T. M.**, (2016, April). A Preliminary Investigation of a Personalized Risk Alert System for Weight Control Lapses. In S.P. Goldstein (Chair), *Harnessing the Power of Predictive Learning to Promote Health Behavior Change: Developing and Testing Novel Technology.* Symposium presented at the 37th Annual Meeting and Scientific Sessions of the Society of Behavioral Medicine, Washington, D.C.

Goldstein, S.P., Evans, B., **Tapera, T. M.**, Forman, E.M., & Butryn, M.L., (2016, April). Is Prompting Problematic?: Considerations for Long-term Ecological Momentary Assessment? Poster presented at the 37th Annual Meeting and Scientific Sessions of the Society of Behavioral Medicine, Washington, D.C.

TEACHING EXPERIENCE

Graduate-level Psychology Statistics Sequence   September 2017 - June 2018
Teaching Assistant   Drexel University
Supervisor: Karol Osipowicz, PhD

- Tutor 1st year Master's students in fundamental statistical procedures relevant to behavioural and social science research:
  - Central tendency and variability,
  - Probability and expectation,
  - Hypothesis testing,
  - Correlation and regression,
  - Chi-square, ANOVA, and variants,
  - Non-parametric tests.
- Direct laboratory sessions to allow students to practice and develop their skills.
- Host office hours to allow students to address lingering questions and concerns.

VOLUNTEER EXPERIENCE

Peer Counseling Helpline   2014 - 2019
Counselor, Vice President of Scheduling & Communications   Drexel University
Supervisor: Scott Sokoloski, PhD

- Chair and record minutes for staff and board meetings.
- Manage ~30 counselors per term and co-ordinate staff shifts and meetings for the term.
- Serve on the board as a counselor liaison.
- Answer calls and counsel students through psychological and emotional distress.
- Analyzed Gmail archive in Python to establish insightful statistics and key performance indicators about the organization.
  - Applied topic modeling algorithms on call logs to classify call topics and caller profiles.

RELEVANT SKILLS AND TOOLS

- Proficient R and Python programming.
- Proficient shell scripting.
- Novice MATLAB, AppleScript, MySQL, Java scripting.
- Qualtrics Survey.
- IRB Certified and Compliant.
- Proficient Microsoft Office Suite.

RELEVANT COURSEWORK

Undergraduate

- Biological Basis of Behavior
- Physiological Psychology
- Neuropsychology
- Industrial/Organizational Psychology
- Experimental Psychology
- Psychological Testing & Assessment
- Dietary Behavior & Technology
- Machine Learning in Psychology
- Introduction to Data Science
- Multivariate Data Analysis
- Introduction to Computer Science
- Programming I & II
- Advanced Programming Techniques & Data Structures
- Mathematical Foundations of Computing

Graduate

- Research Methods I & II
- Cognitive Psychology
- Citizen Science
- Behaviour Analysis
- Statistical Data Analysis I&II
- Statistical Data Analysis III: Advanced Topics
- Behavioural Data Mining
- Hierarchical modelling
- Theoretical Computer Science Foundations
- Programming Foundations
- Advanced Programming Techniques

Please visit https://github.com/TinasheMTapera for code samples and projects.

REFERENCES

Theodore Satterthwaite, PhD
Principal Investigator, Director
Penn Lifespan Informatics & Neuroimaging Center
University of Pennsylvania
3700 Hamilton Walk, Richards 5A
Philadelphia, PA 19104
Phone: 267-441-7145
Email: sattertt@pennmedicine.upenn.edu

Matthew Cieslak, PhD
Senior Data Scientist
Penn Lifespan Informatics & Neuroimaging Center
University of Pennsylvania
3700 Hamilton Walk, Richards 5A
Philadelphia, PA 19104
Email: Matthew.Cieslak@pennmedicine.upenn.edu

Fengqing Zoe Zhang, PhD
Assistant Professor
MS Research Mentor
Drexel University
3141 Chestnut Street, Stratton 326
Philadelphia, PA 19104
Phone: 215.553.7172
Email: fengqing.zhang@drexel.edu